THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Systems Biology of Protein Synthesis and Secretion in Yeast

QI QI

Department of Biology and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021

Systems Biology of Protein Synthesis and Secretion in Yeast

QI QI

ISBN 978-91-7905-556-1

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5023
ISSN 0346-718X

Division of Systems and Synthetic Biology
Department of Biology and Biological Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover: The Design-Build-Test-Learn (DBTL) workflow for investigation of the protein synthesis and secretion process in yeast.

Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

Systems biology of protein synthesis and secretion in yeast

Qi Qi
Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

Protein synthesis and secretion is a vital process to maintain cell function. As it demands numerous building blocks, cofactors and chaperones generated from metabolism and translation, the process is intertwined with metabolic and regulatory networks. To obtain an overall understanding of the protein synthesis and secretory system, multi-omics data are coupled with mathematical modeling to systematically quantify cellular resource reallocation in response to recombinant protein production and/or nutrient starvation.

In this thesis, we mainly use two recombinant proteins, α-amylase and insulin precursor, as model proteins to study the protein synthesis and secretion process in a model organism *Saccharomyces cerevisiae*. We find that the central metabolism is reprogrammed at a large scale to relieve the oxidative stress caused by recombinant protein production, and the activation of Gcn2p-mediated signaling pathway plays a crucial role in reshaping metabolism. As protein folding is often considered the flux controlling step in protein synthesis and secretion, we further identify two routes of the protein folding pathway to improve protein production, namely through improved folding capacity and increased folding precision, respectively. Additionally, protein translation is the initial step of protein synthesis. We find that cells maintain large and unequally distributed reserves in translational capacity by stepwise reducing nitrogen availability. Moreover, we also construct a proteome-constrained genome-scale protein secretory model for *S. cerevisiae* (pcSecYeast) to perform secretory simulations and provide genomic targets for cell engineering. Our findings elucidate the global responses to various perturbations on protein synthesis and secretion and provide valuable novel insights that can be leveraged for improving recombinant protein production.

Key words: protein synthesis, protein secretion, multi-omics analysis, genome-scale modeling, recombinant protein production

List of Publications

This thesis is based on the work contained in the following publications and manuscripts:

Paper I: **Qi Qi**, Feiran Li, Egor Vorontsov, Jens Nielsen. Recombinant protein production requires metabolic reprogramming to provide more NADPH via activation of kinase Gcn2p. [Under review]

Paper II: **Qi Qi**, Feiran Li, Rosemary Yu, Martin Engqvist, Verena Siewers, Johannes Fuchs, Jens Nielsen. Different routes of protein folding contribute to improved protein production in *Saccharomyces cerevisiae*[J]. *mBio*, 2020, 11(6).

Paper III: Rosemary Yu, Kate Campbell, Rui Pereira, Johan Björkeroth, **Qi Qi**, Egor Vorontsov, Carina Sihlbom, Jens Nielsen. Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast[J]. *Nature communications*, 2020, 11(1): 1-12.

Paper IV: Feiran Li, Yu Chen*, **Qi Qi***, Yanyan Wang*, Le Yuan, Ibrahim EI-Semman, Amir Feizi, Eduard Kerkhoven, Jens Nielsen. Genome-scale modeling of the protein secretory pathway reveals novel targets for improved recombinant protein production in yeast. [Manuscript]

Additional publications not included in the thesis:

Paper V: Yuping Lin*, Hongzhong Lu*, Yufeng Guo*, **Qi Qi***, Jun Geng*, Feiran Li, Qianni Qi, Yuanyuan Zhang, Jens Nielsen, Qinhong Wang, Yanhe Ma. Genome-wide single-nucleotide evolution endowing *Saccharomyces cerevisiae* with stress tolerance. [Manuscript]

Paper VI: John Hellgren, **Qi Qi**, Jens Nielsen, Verena Siewers. Proteome and transcriptome analysis of a yeast strain expressing a cyclic phosphoketolase pathway for improved acetyl-CoA supply. [Manuscript]

*** Contributed equally**

Contribution Summary

Paper I: Designed the study, performed the experiments, analyzed the data and wrote the manuscript

Paper II: Designed the study, performed the experiments, analyzed the data and wrote the manuscript

Paper III: Performed the experiments and assisted in manuscript preparation

Paper IV: Performed the experiments and assisted in manuscript preparation

Paper V: Co-designed the study, performed the experiments, analyzed the data and assisted in manuscript preparation

Paper VI: Performed the experiments and assisted in manuscript preparation

Preface

This dissertation serves as partial fulfillment of the requirements to obtain the degree of Doctor of Philosophy at the Department of Biology and Biological Engineering at Chalmers University of Technology. The PhD studies were carried out between September 2017 and October 2021 at the division of Systems and Synthetic Biology under the supervision of Professor Jens Nielsen. The project was mainly funded by the Novo Nordisk Foundation and the Swedish Foundation for Strategic Research.

Qi Qi
October 2021

Table of Contents

Abbreviations

ATP	adenosine triphosphate
COPII	coat protein complex II
DCW	dry cell weight
DNA	deoxyribonucleic acid
ER	endoplasmic reticulum
ERAD	ER-associated degradation
FBA	flux balance analysis
FSEOF	flux scanning based on enforced objective function
GEM	genome-scale metabolic model
GPI	glycosylphosphatidylinositol
IP	insulin precursor
mRNA	messenger ribonucleic acid
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NGAM	non-growth associated maintenance energy
PPP	pentose phosphate pathway
PTM	post-translational modification
TCA	tricarboxylic acid
TORC1	target of rapamycin complex 1
tRNA	transfer ribonucleic acid
UDP	uridine diphosphate
UPR	unfolded protein response

I don’t know anything, but I do know that
everything is interesting if you go into it deeply enough.

Richard P. Feynman

1. Introduction

1.1 *Saccharomyces cerevisiae*

The budding yeast *Saccharomyces cerevisiae*, which has been used for alcoholic fermentation for thousands of years (1, 2), is one of the most popular eukaryal organisms for studies of cell systems and production of various chemicals and recombinant proteins (3–8). The popularity benefits by many properties *S. cerevisiae* possesses, including i) its complex cellular structure contributes to studies of eukaryal molecular mechanisms; ii) high growth rate compared to other eukaryal organisms; iii) much information available about this organism through high-throughput studies, databases, sequenced genomes and extensive toolbox for molecular modification.

1.2 Protein synthesis and secretion

The central dogma of molecular biology states the flow of genetic information within a biological system. The transcription and translation processes initiate the protein synthesis process in cells. mRNA guides the assembly of free amino acids to synthesize polypeptide in ribosome with the assistance from tRNA. Next, the nascent polypeptide needs to be processed in the endoplasmic reticulum (ER) and the Golgi to form native protein structure, with help of the protein secretory pathway.

Protein secretion is defined as the process of maturation of proteins destined for secretion, for the plasma membrane, the organelles or the extracellular space. In the 1960s, along with the development of electron microscopy, George Palade established the protein secretory pathway and its connection to the biogenesis of organelles (9). The protein secretory pathway in eukaryal cells is an elaborate machinery that consists of a number of independent organelles that work together efficiently to secret proteins to different destinations (10–13). Each organelle provides a specialized environment that facilitates the various stages in protein biogenesis, modification, folding and sorting. In yeast this pathway involves more than 160 proteins that are responsible for different post-translation processes (14). Most of these proteins are located in the ER, which is an important compartment in the secretory pathway.

Protein translocation

The secretory pathway initiates from translocation of newly synthesized polypeptide through ER membrane. The translocation process can occur either co-translationally where translation and translocation are directly coupled, or post-translationally depending on the hydrophobicity and amino acid composition of the already fully translated signal peptide (15, 16). In this process, the Sec61 complex is a commonly used channel for protein import into the ER.

Protein glycosylation

Since the majority of all secreted proteins are glycoproteins, the *N*-glycosylation is one of the most prevalent post-translational modifications in the ER. The *N*-linked oligosaccharides play an essential role in the protein folding quality control system. In the beginning, an oligosaccharide precursor $Glc_3Man_9GlcNAc_2$ is assembled on the asparagine residue of the *N*-glycosylation recognition sequence of the nascent polypeptide during translocation (17), assisted by the ER-resident oligosaccharyltransferase. Additionally, at the hydroxyl groups of serine and threonine there will be *O*-glycosylation modification (18), which is catalyzed by protein *O*-mannosyltransferase.

The protein folding quality control system

The term 'quality control' describes the process of conformation-dependent molecular sorting of newly synthesized proteins in the ER (19). During the protein folding process, besides correctly folded proteins, there is also a fraction of misfolded and incompletely assembled proteins.

In yeast, the secretory pathway is responsible for the modification and maturation of more than 550 proteins which are vital to maintain cell function. High-fidelity protein folding is a crucial step in the secretory pathway. Misfolded proteins, which can be caused by protein overproduction, externally applied stresses such as heat or oxidative stress, or genetic factors such as genetic mutations, transcription errors, or translation errors, are handled by the quality control system to be either refolded or delivered to the proteasome for degradation through ER-associated degradation (ERAD) pathway (20–22). Without proper handling by this quality control system, accumulation of misfolded proteins can be toxic to the cell. In humans, such accumulation can cause a number of severe diseases, such as Alzheimer's disease, Parkinson's disease, and type II diabetes (23). As a model eukaryal organism, the yeast *S.*

cerevisiae is a good model to study the secretory pathway and unravel the mechanisms related to these diseases (24).

Unfolded protein response

The accumulation of misfolded proteins results in ER stress that is handled by the unfolded protein response (UPR). Activation of the UPR results in transcriptional change of a large number of genes, which relates to around 400 genes in yeast (25). A result of this regulation is upregulation of chaperones and foldases as well as ERAD (26–28). Based on studies of the UPR many targets for improving protein secretion have been identified and implemented (7), and it is generally believed that any factor that reduces ER stress and its downstream damage caused by recombinant protein production results in improved secretion of the produced protein, and it is therefore important to engineer the secretory pathway to have the right capacity to process the protein of interest.

Post ER processing

The correctly folded proteins are transported into the Golgi for further processing. The coat protein complex II (COPII)-coated vesicles mediate the trafficking from the ER to the Golgi (29). In the Golgi, proteins are further maturated, e.g. hyper-mannosylation. After processing, proteins are sorted to different destinations, such as endosome, vacuole, plasma membrane or extracellular space, depending on their properties. Noteworthy, retrograde transport to the ER is also occurred to return cargo adaptor proteins, membrane components and escaped ER resident proteins, which is mediated by the coat protein complex I (COPI) (30, 31).

1.3 The approaches in systems biology

With the development of genomics, omics analysis and detailed mathematical models in the field of systems biology, it has become possible to perform very detailed phenotypic characterization of cells and use these for guiding metabolic engineering (4, 32). There are two different approaches to systems biology: in top-down systems biology, different types of high-throughput generated data, often referred to as omics data, are analyzed in an integrative approach, and in bottom-up systems biology, detailed models for specific processes, for example, enzymatic reactions, are assembled into a model describing the system being studied. The two approaches are complementary; top-down systems biology is useful for mapping cellular functions at the genome scale, whereas bottom-up systems biology enables detailed timescale resolution of the impact of individual components on overall system properties.

Integrative analysis of omics data

The availability of high-throughput experimental techniques, often referred to as omics techniques, has allowed a more in-depth study of cell systems. In particular, it has become possible to begin to address the general question of how cellular resources are allocated through integrative analysis of transcriptome, proteome, phosphoproteome, metabolome and fluxome. Integrative analysis of omics data usually uses biological networks to make it possible to identify parts of large networks that are coregulated (33). In this context, metabolic networks are therefore well suited because they are reconstructed using detailed biochemical information (34). Other biological networks, including gene ontology (GO) annotations that also reveal connections between genes, can also be used to identify gene-enrichment groups or reporter features (34, 35), for example, reporter metabolites or reporter transcription factors.

In this thesis, we study the cell systems through integrative study of transcriptome, proteome, phosphoproteome and exo-metabolome, and the analysis of GO bioprocess enrichment provides much useful information.

Mathematical modeling

For describing metabolism, mathematical models can be divided into two groups: kinetic models and stoichiometric models. In kinetic models, reaction rates are modeled as a function of metabolite concentrations and metabolite concentrations are modeled as a function of time (36). The model construction needs input of plenty of parameters, making the typical size of the model in a small scale (<100 metabolites). In stoichiometric models, a pseudo-steady state is assumed, and therefore, metabolite concentrations are not modeled and reaction rates are inferred by imposing steady-state mass balances on each metabolite (37), which makes stoichiometric models easier to be constructed in a large scale, even in the genome scale.

In this thesis, we only focus on stoichiometric models. The genome-scale metabolic model (GEM) in yeast (Yeast8) (38) is used to simulate most flux distribution and enzyme usage in cells. In addition, we also construct a proteome-constrained secretory model (pcSecYeast) based on Yeast8, to show the particular molecular mechanisms in the secretory pathway.

1.4 Recombinant protein production

The production of recombinant proteins by microorganisms or cell cultures, including biopharmaceuticals and industrial enzymes, is a growing multibillion-dollar industry (39). Among these recombinant proteins, about 20% are being produced by *S. cerevisiae* (40, 41).

The advantages of using the yeast *S. cerevisiae* as a cell factory for the production of recombinant proteins are that the proteins can be secreted to the extracellular medium and this facilitates subsequent purification, and in many cases yeast can perform proper post-translational modifications of the proteins, including proteolytic processing of signal peptides, disulfide bond formation, subunit assembly, acylation and glycosylation.

In this thesis, we study the synthesis and secretion of two model recombinant proteins, insulin precursor (IP) and α-amylase. IP is a single chain peptide, containing 53 amino acid residues. There are three disulfide bonds and no *N*-glycosylation site in its native structure. α-Amylase contains 478 amino acid residues. There are three domains, four disulfide bonds and one *N*-glycosylation site in its native structure. Investigation of these two proteins, with distinct molecular size, post-translational modifications and three-dimensional structures, makes it possible to obtain a universal understanding of the protein synthesis and secretion process.

1.5 Chemostat culture

Chemostat culture is one kind of typical modes of continuous cultivation. In chemostat, the medium is designed to ensure that there is a single limiting substrate, which allows for controlled variation in the specific growth rate of the biomass (42). For *S. cerevisiae* fermentation on minimal mineral medium, glucose is usually the limiting substrate in carbon-limited chemostat culture and $(NH_4)_2SO_4$ is usually the limiting substrate in nitrogen-limited chemostat culture. Since in a steady-state continuous bioreactor, the specific growth rate equals the dilution rate, the cells can be cultivated at different specific growth rates through control of dilution rate (or feed flow rate), which allows for precise experimental determination of the specific rates and is attractive for physiological studies.

However, there is an upper threshold of dilution rate. Based on the Monod model, the mass balance of the biomass gives:

$$D = \mu_{max} \frac{c_s}{c_s + K_s}$$

Here D is referred to as the dilution rate, μ_{max} as the maximum specific growth rate of the cell, c_s as the concentration of the limited substrate, and K_s as a specific substrate-related parameter. The value of K_s equals the substrate concentration at which the specific growth rate is 0.5 μ_{max}. Since for a specific cell, the μ_{max} and K_s are constant, it is easier to know that D and c_s keep positive correlation, and when c_s increase to the substrate concentration

in the feed c_s^f, the dilution rate attains its maximum value, which is called the critical dilution rate:

$$D_{crit} = \mu_{max} \frac{c_s^f}{c_s^f + K_s}$$

When the dilution rate becomes equal to or larger than D_{crit}, the biomass is washed out of the bioreactor. A commonly used empirical value of D_{crit} for *S. cerevisiae* is 0.8-fold μ_{max}.

In addition to the advantages of precise control of specific growth rate and collection of physiological data at the steady state, chemostat can also be used to study dynamic conditions. For example, the substrate concentration can be suddenly changed by adding a pulse of the limiting substrate to the bioreactor or the temperature can be changed by adjustment of the heating system, and the cellular responses to the environmental changes can be studied.

As mentioned above, chemostat culture can be used to avoid bias caused by differences in the growth rates of different strains and delays in protein synthesis, this mode is generally performed to study the physiological differences between different strains. In addition, for industrial application, although the product concentration in chemostat is often lower than the titer in fed-batch culture, the productivity could be high because of the high efficiency of continuous fermentation. It is reported that chemostat culture has been used for recombinant protein production in an industrial setting (43).

In this thesis, all omics data are collected from cells grown in steady-state chemostat cultures.

1.6 Aims and significance

In this thesis, we study the protein synthesis and secretion process in *S. cerevisiae* through combination of integrative analysis of multi-omics data and mathematical modeling.

Systematic characterization of yeast strains with recombinant protein production is a feasible approach to comprehensively understand the protein synthesis and secretion process. Firstly, we performed orthogonal experiments to evaluate strains producing proteins of different complexities, and at different yields, compared with the reference strain (*from zero to one*) (**Paper I**). We found that recombinant protein production remodeled the central carbon metabolism at a large scale, including decreased fermentation and respiration, enhanced glycolysis and tricarboxylic acid cycle (TCA cycle) to supply more cytosolic nicotinamide adenine dinucleotidephosphate (NADPH) and precursors for amino acids biosynthesis, and

increased glycerol production to maintain cellular redox potential. Finally, by analyzing phosphoproteomic abundance, we demonstrated a key role of Gcn2p-mediated signaling pathway on reshaping metabolism. Next, we investigated the rate-limiting steps in the secretory pathway for improving recombinant protein production, through characterization of evolved strains with high recombinant protein productivity, compared with the original strain (*from one to hundreds*) (**Paper II**). We identified different routes of the protein folding pathway to improve protein production, namely through improved folding capacity and increased folding precision, respectively. We further found that the regulation of *N*-glycans played an important role in the folding precision control, and overexpression of the glucosidase Cwh41p can significantly improve the protein production.

Protein translation is the initial step of protein synthesis and secretion. As nitrogen is an important component of nucleotide and amino acid, herein we investigated the translation propensity through examining transcriptome, proteome and ribosomal protein abundance of yeast strains with different nitrogen accessibility (**Paper III**). We found that yeast growing in typical laboratory conditions maintains an overall 50% reserve proteome, 75% reserve transcriptome, and 50% reserve translational capacity, and ribosome reserves contained up to 30% sub-stoichiometric ribosomal proteins, with activation of reserve translational capacity associated with selective upregulation of 17 ribosomal proteins. We also identified that a major part of the translation capacity reserves is preferentially maintained for metabolic processes, highlighting the importance of a robust metabolism for cell growth and survival.

Development of a secretory pathway-contained metabolic model contributes to comprehensive understanding of the process. In the final part of the thesis, I introduce our work on the construction of pcSecYeast (**Paper IV**). We added detailed reactions related to the secretory pathway, including translocation, glycosylation, folding and degradation, into Yeast8 to construct the model. We simulated the effects of protein misfolding on cell growth and the recombinant protein production process. We further predicted genomic targets for cell engineering to improve recombinant protein production with high accuracy.

The studies in this thesis also demonstrate that how integrative analysis of omics data and mathematical modeling can be leveraged to provide valuable novel insights for systems biology and metabolic engineering.

2. The protein secretory pathway

2.1 The link between protein secretory pathway and cellular central metabolism

As mentioned previously, in this section, the original strain Qc, which carries an empty plasmid, was used as the reference strain, and four strains AIP, AIC, AAP and AAC, which carry a plasmid with recombinant protein production cassette but with different promoters (*TEF1p* or *TPI1p*) or producing different proteins (IP or α-amylase) were studied. Therefore, the genomes of all strains used in this section only are different on the promoters controlling the expression of recombinant protein production (Fig. 2-1), which helps eliminate the differences between strains caused by genome mutation or strain engineering. The strains were grown under chemostat culture with a specific growth rate of 0.2/h to further eliminate the differences on growth rate. We examined physiological parameters, proteome, phosphoproteome, transcriptome and intracellular amino acid concentrations of these five strains to elucidate the global cellular responses to recombinant protein production.

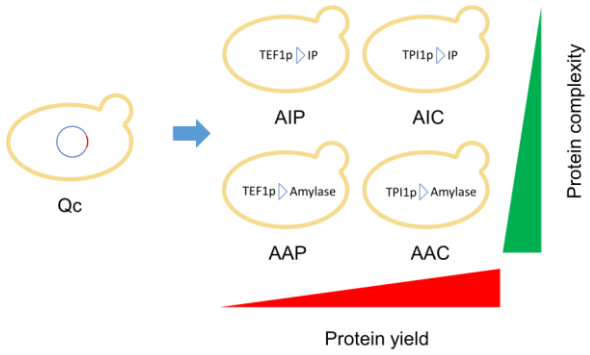

Figure 2-1. Strains used in Paper I. IP, insulin precursor; Amylase, α-amylase.

In chemostat cultures, the *TPI1* promoter was found to be superior over the *TEF1* promoter for production of both IP and α-amylase (Fig. 2-2a), which is in line with previous results in batch cultures (44). The protein production strains exhibited decreased ethanol production (Fig. 2-2b) and increased glycerol production (Fig. 2-2c) compared to the reference strain Qc, with larger differences in high yield strains (AIC and AAC), while all five strains exhibited comparable glucose uptake yields (Fig. 2-2d), biomass (Fig. 2-2e) and acetate production yields (Fig. 2-2f). The decreased ethanol yield indicates reduced fermentation and potentially

enhanced TCA cycle in the protein production strains. The increased glycerol yield suggests the importance of redox capacity for protein production, as the essential role of glycerol production in the maintenance of the redox potential (45). Furthermore, of the 14 amino acids supplemented in the media, several were taken up at higher rates by the protein production strains, including methionine, phenylalanine, tyrosine, tryptophan, leucine and threonine (Fig. 2-2g). Uptake of methionine contributes to maintenance of redox balancing, as biosynthesis of one molecule of methionine *in vivo* requires three molecules of NADPH (46). In addition, the biosynthesis of three aromatic amino acids, phenylalanine, tyrosine and tryptophan, as well as leucine, demands more energy than others (47), indicating increased uptake of amino acids is driven by an economical strategy. All other amino acids showed same uptake rates between the strains.

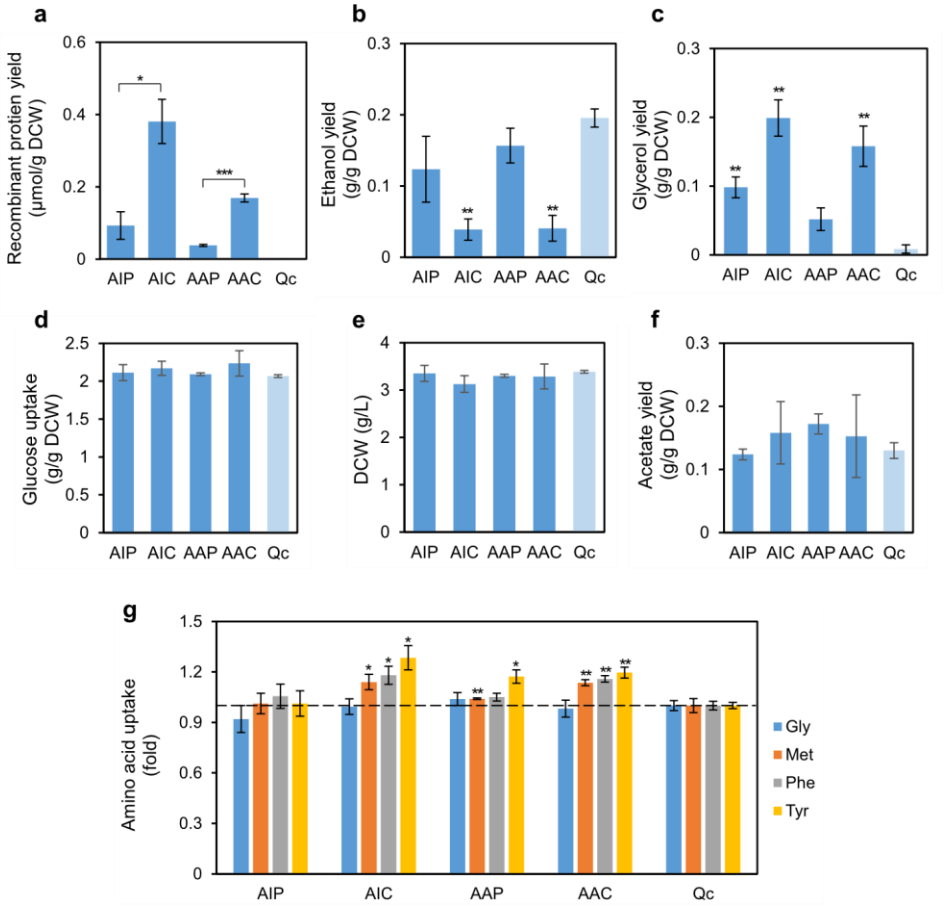

Figure 2-2. Overview of physiological parameters between strains. a Recombinant protein yield for each relevant strain. Insulin precursor for AIP and AIC; α-amylase for AAP and AAC. (**b** and **c**) Ethanol and glycerol yield for each strain. **d** Glucose uptake for each strain. **e** Final biomass for each strain. **f** Acetate yield for each strain. **g** Changes of uptake yield of glycine, methionine, phenylalanine and tyrosine in the protein production strains compared with Qc. Data shown are mean values ± standard errors of the means of biological triplicates. Statistical significance was determined by a two-tailed Student's *t* test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

We obtained absolute quantities of 3 944 proteins for each strain by proteome analysis. We further calculated the proteome allocation of 99 Yeast GO-Slim bioprocesses based on GO Slim Mapper annotations (48), and found that a number of processes were significantly re-allocated in all protein production strains compared to the Qc strain ($P < 0.05$) (Fig. 2-3a). Several processes related to central carbon metabolism was remodeled, in either direction, in response to recombinant protein production: allocation to metabolic energy generation, carbohydrate metabolic process and monocarboxylic acid metabolic process were increased, but allocation to cellular respiration, mitochondrial translation and mitochondrion organization were decreased (Fig. 2-3a). Additionally, proteome allocation to several processes relevant to protein secretion were increased, including membrane fusion, organelle fusion and vacuole organization, while carbohydrate transport and ion transport were decreased.

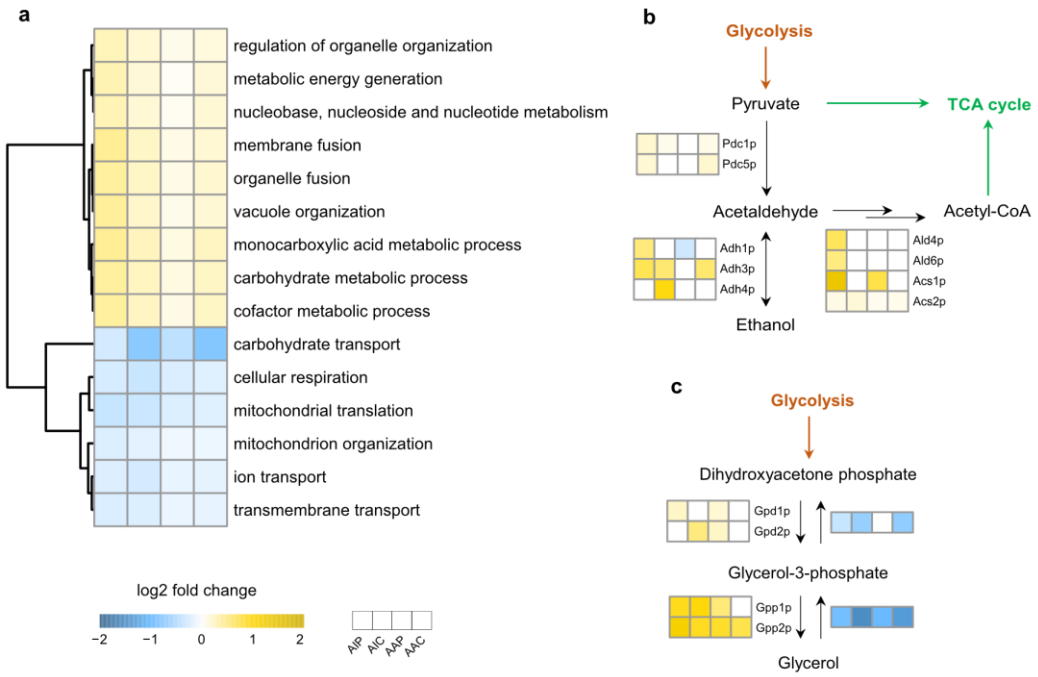

Figure 2-3. Proteome analysis reveals the changes of protein expression in response to recombinant protein production. a The differentially expressed biological processes ($P < 0.05$) at the proteome level in all protein production strains. All proteins were allocated to the 99 Yeast GO-Slim biological processes. **b** Differentially expressed proteins ($P < 0.05$) relevant to ethanol production. **c** Differentially expressed proteins ($P < 0.05$) relevant to glycerol production. The expression levels of processes or proteins in Qc were used as the references.

Ethanol and glycerol are the two primary by-products of central carbon metabolism in yeast. For ethanol metabolism, increased abundance of Ald4p, Ald6p, Acs1p and Acs2p in protein production strains suggested that more acetaldehyde was converted to Acetyl-CoA, which also induced the enhanced ethanol catabolism (Fig. 2-3b), in accord with the decreased

ethanol yield in protein production strains (Fig. 2-2b). Random sampling analysis also showed that the pyruvate to Acetyl-CoA reaction carried more flux in the protein production strains than Qc, while the pyruvate to acetaldehyde reaction carried less flux (Fig. 2-4). For glycerol metabolism, in line with the increased glycerol yield in the protein production strains (Fig. 2-2c), enzymes in its anabolic pathway were upregulated and enzymes in its catabolic pathway were downregulated (Fig. 2-3c).

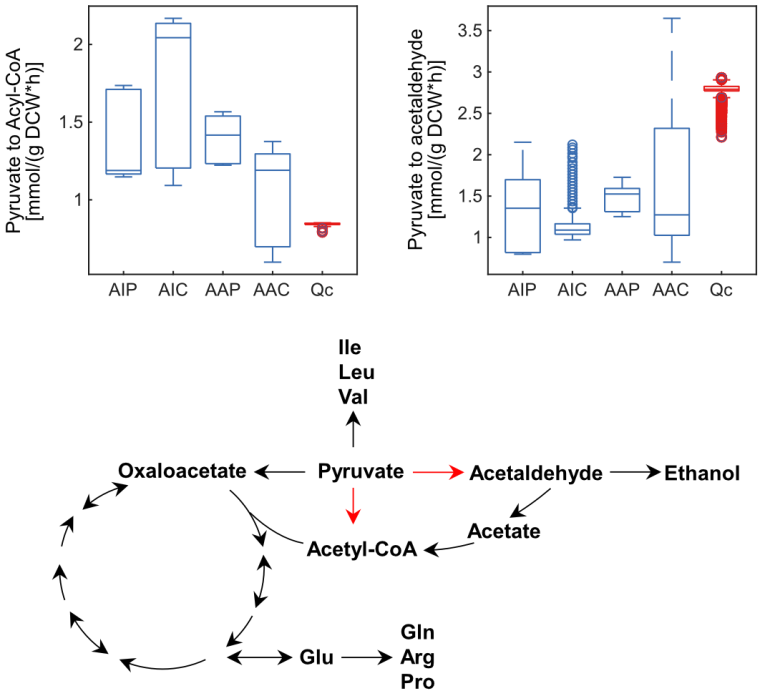

Figure 2-4. Carriable flux of the reaction pyruvate to Acyl-CoA or pyruvate to acetaldehyde in each strain.

To further investigate central carbon metabolism, we compared the proteome allocation to glycolysis and the pentose phosphate pathway (PPP) between the protein production strains and Qc. In Qc, there was 7.9% of the proteome allocated to glycolysis, and to produce recombinant proteins, the allocation increased to 8.9% - 12.3% (Fig. 2-5a). On the other hand, the PPP allocation remained the same, at 0.6% in Qc, and 0.5% - 0.7% in the protein production strains (Fig. 2-5b).

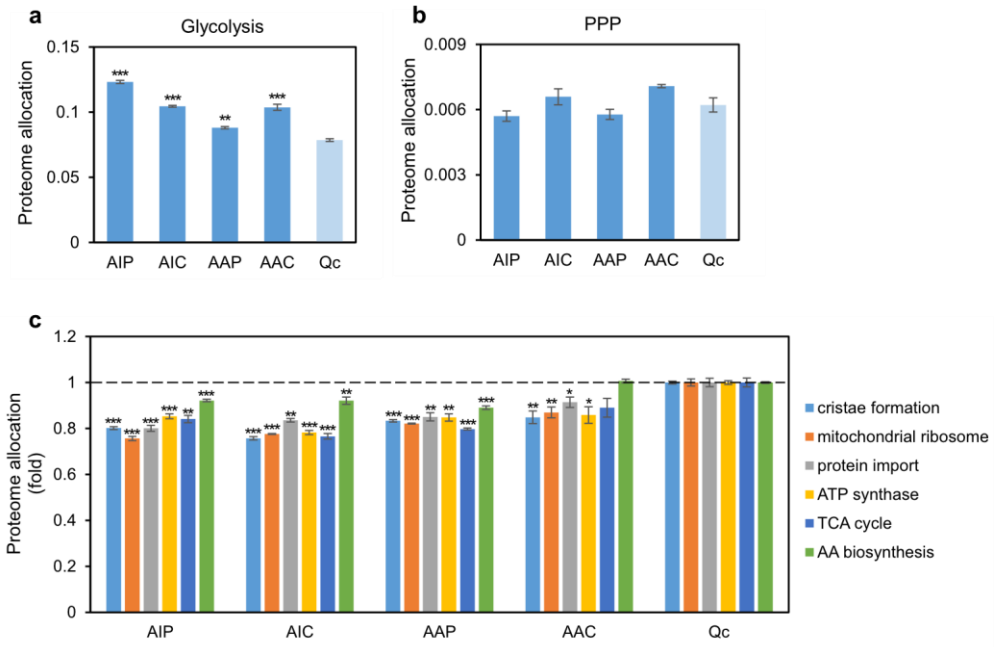

Figure 2-5. Proteome reallocation reveals the changes of central carbon metabolism in response to recombinant protein production. a Total proteome allocation to all enzymes in glycolysis was compared between protein production strains and Qc. **b** Total proteome allocation to all enzymes in the pentose phosphate pathway (PPP) was compared between protein production strains and Qc. **c** Changes of mitochondrial subgroups in the protein production strains compared with Qc at the proteome level. Data shown are mean values ± standard errors of the means of biological triplicates. Statistical significance was determined by a two-tailed Student's *t* test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Proteomics analysis identified many downregulated processes present in the mitochondrion (Fig. 2-2a). To further unravel the mitochondrial changes in response to recombinant protein production, proteins located in the mitochondrion are allocated to different subgroups based on mitochondrial gene ontology (49) and significantly changed subgroups are shown in Fig. 2-5c. Compared with Qc, most mitochondrial components had significantly decreased abundance in the protein production strains. In particular, the abundance of ATP synthases is significantly decreased in the protein production strains, indicating attenuated respiration. Correspondingly, flux balance analysis (FBA) results also showed that the relative ATP yield on biomass generated from respiration in these strains decreased compared with Qc (3.2%, 8.6%, 8.2% and 1.1% decrease for AIP, AIC, AAP and AAC, respectively) (Fig. 2-6a). Conversely, the ATP yield from the TCA cycle increased to meet the cellular energy demand (43.9%, 88.2%, 47.7% and 20.2% increase for AIP, AIC, AAP and AAC, respectively) (Fig. 2-6a). Moreover, while the abundance of enzymes in the TCA cycle decreased in the protein production strains, FBA results revealed that most enzymes exhibited a greater usage ratio and generally carried higher flux than the enzymes in Qc. Similar observations were made

for enzymes in amino acid biosynthesis as well. Further comparison of the fluxes showed that fluxes in respiration were comparable (Fig. 2-6b), but the fluxes with ATP generation in the TCA cycle increased a lot in the protein production strains (Fig. 2-6c). Hence, the total ATP yield either did not decrease (in AIP, AAP and AAC) or only slightly decreased (3.9% decreased in AIC).

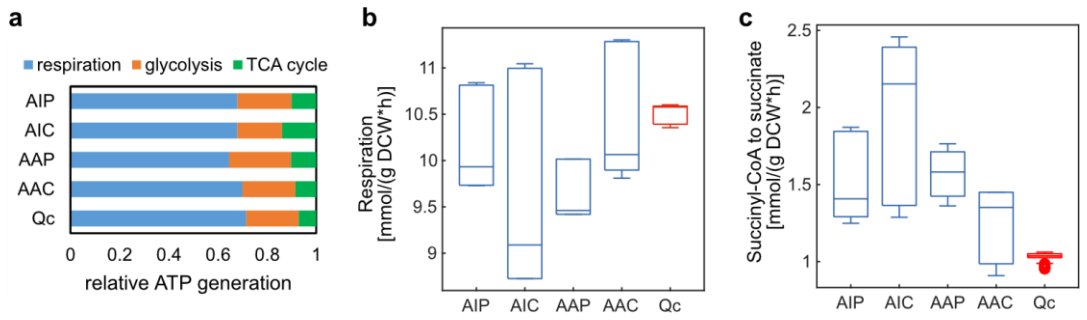

Figure 2-6. Changes in the energy generation pathways. a Relative ATP yield generated from respiration, glycolysis and TCA cycle between strains. **b** Carriable flux of ATP synthase in the mitochondrion. **c** Carriable flux of succinyl-CoA to succinate in the TCA cycle.

The TCA cycle supplies the reducing agent nicotinamide adenine dinucleotide (NADH) for the respiratory chain resulting in ATP generation. Compared with Qc, the enhanced TCA cycle flux and decreased respiration in the protein production strains indicated an imbalance in NADH formation and consumption. Further investigation, however, showed that the enhanced TCA cycle generated more cytosolic NADPH and not mitochondrial NADH through oxidation of isocitrate to α-ketoglutarate (AKG). The oxidation of isocitrate to AKG in the mitochondrion can be catalyzed by Idh1p, Idh2p, or Idp1p, producing either NADH (Idh1p and Idh2p) or NADPH (Idp1p), and in the cytoplasm, this reaction is catalyzed by Idp2p only and is NADP-specific (50) (Fig. 2-7a). Proteome analysis revealed that with same total amount of these four enzymes, the fraction of Idp2p increased in the protein production strains (Fig. 2-7b). In line with this, random sampling analysis showed that the mitochondrial NAD-specific isocitrate dehydrogenative reaction carried comparable flux in each strain (Fig. 2-7c), but the cytosolic NADP-specific reaction in protein production strains was able to carry 7.9-fold to 13.1-fold higher flux than in Qc (maximum carriable flux was 1.57, 1.10, 1.32, 0.95 and 0.12 for AIP, AIC, AAP, AAC and Qc, respectively, with the unit of mmol $gDCW^{-1}$ h^{-1}) (Fig. 2-7d). At the same time, the flux carried by the mitochondrial NADP-specific reaction was very low in each strain (Fig. 2-7e), indicating a low NADPH demand in the mitochondrion. The PPP is also a crucial pathway to generate cytosolic NADPH, but in the protein production strains, the two NADPH-generated reactions in the PPP carried low

and comparable fluxes, compared with Qc (Fig. 2-7f and g), in agreement with the result that a similar fraction of proteome was allocated to the PPP in each strain (Fig. 2-5b).

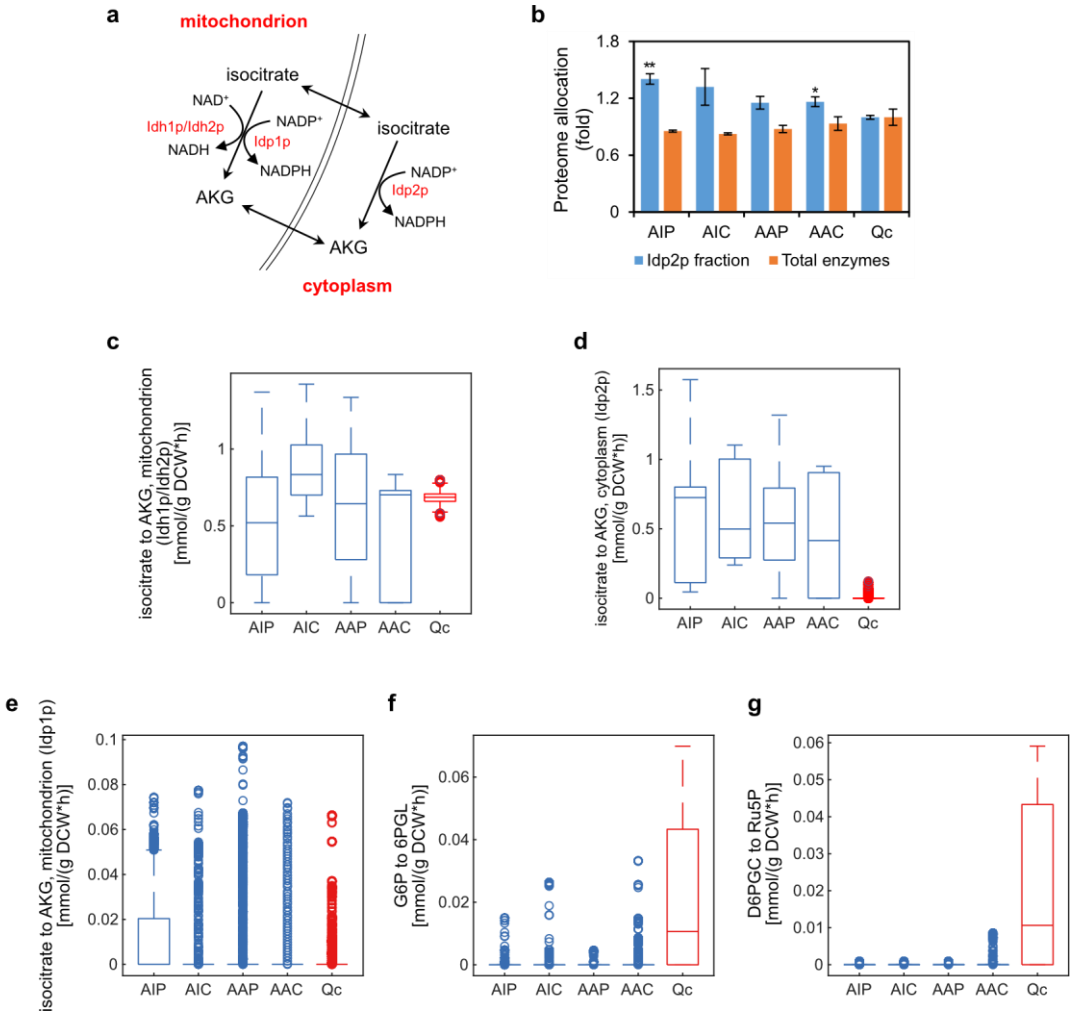

Figure 2-7. More cytosolic NADPH is provided by oxidation of isocitrate to AKG. a Oxidation of isocitrate to AKG is under catalysis of four enzymes. **b** Proteome allocation to Idp2p and total four enzymes catalyzing the oxidation of isocitrate to AKG. **c** Carriable flux of isocitrate to AKG catalyzed by Idh1p and Idh2p. **d** Carriable flux of isocitrate to AKG catalyzed by Idp2p. **e** Carrible flux of isocitrate to AKG catalyzed by Idp1p in the TCA cycle. **f** Carriable flux of G6P to 6PGL in the PPP. **g** Carriable flux of D6PGC to Ru5P in the PPP. AKG, α-ketoglutarate. Carriable fluxes of relevant reactions with NADH or NADPH generation. G6P, glucose 6-phosphate; 6PGL, 6-phosphogluconolactonase; D6PGC, 6-phosphogluconate; Ru5P, ribulose-5-phosphate. For panel **b**, data shown are mean values ± standard errors of the means of biological triplicates. Statistical significance was determined by a two-tailed Student's *t* test. *, $P < 0.05$; **, $P < 0.01$.

In phosphoproteomic analyses, we identified 836-1,533 phosphorylated sites that are differentially expressed, in any one of the protein production strains, compared to Qc ($P < 0.05$) (Fig. 2-8a). Regarding the overlapped 140 phosphorylated sites which were significantly changed, the abundance of 25 sites was upregulated and the abundance of 110 sites was downregulated in all protein production strains (Fig. 2-8b). The highly conserved

regulation of phosphorylation events (135 out of 140 sites) indicated the importance of these proteins on protein production.

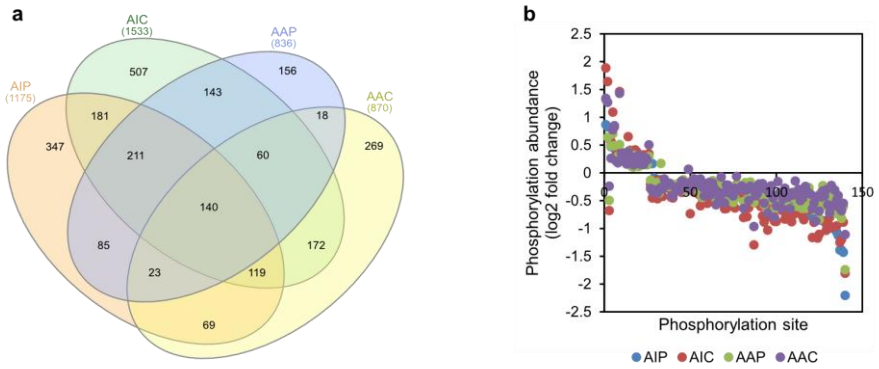

Figure 2-8. Overview of phosphoproteomics changes. a Overlaps of phosphorylation sites with significantly changed abundance ($P < 0.05$) compared with Qc between protein production strains. **b** Changes of phosphorylation abundance regarding the overlapped 140 protein sites.

GO term enrichment analysis of these proteins revealed several processes relevant to translation, including regulation of translational initiation, mature ribosome assembly and translational initiation. At the protein level, there were 6 proteins related to translation, i.e. Fun12p, Gcd6p, Gcn2p, Gcd1p, Tif5p and Tif6p. The phosphorylated forms of these 6 proteins were significantly decreased in the protein production strains compared with Qc (Fig. 2-9a), but the abundance of these proteins did not show any differences (Fig. 2-9b), indicating the importance of protein phosphorylation as a regulatory mechanism of translation-related processes.

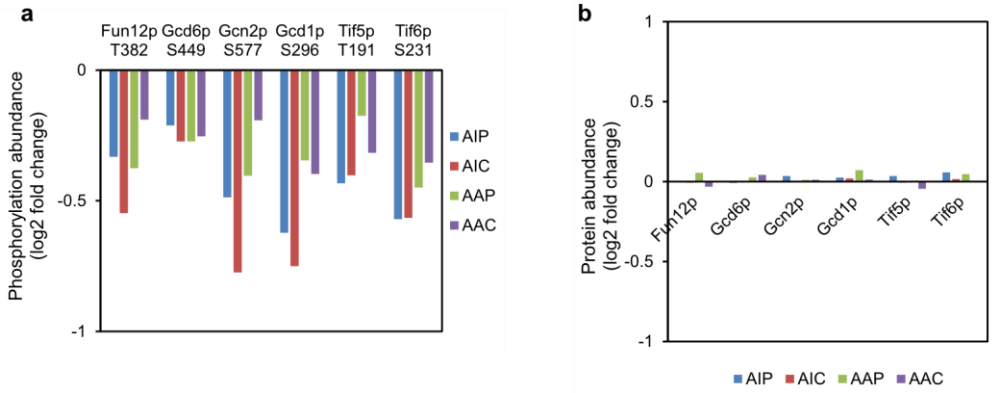

Figure 2-9. Overview of phosphoproteomics and proteomics changes of proteins involved in translation process. a Changes of phosphorylation abundance in proteins relevant to translational regulation compared with Qc. **b** Changes of abundance at the protein level compared with Qc.

Of these 6 proteins, Gcn2p is an important protein kinase responsible for sensing starvation in a dynamic environment (51). Gcn2p is activated via dephosphorylation of Gcn2p-Ser-577 in amino acids-deprived cells (52), which was observed in all protein production strains (Fig. 2-9a). The activated Gcn2p is able to phosphorylate the α subunit of eukaryotic initiation factor 2 (eIF2α) and further reduce general protein biosynthesis but specifically induce the expression of Gcn4p and over 500 genes under its control (Fig. 2-10a), including many amino acid biosynthetic genes (53). Since amino acid starvation is known to be a common signal to activate the Gcn2p signaling pathway (54), we therefore examined the intracellular concentrations of 20 natural amino acids for each strain. Our results showed that, overall there were less free amino acids in the protein production strains than Qc, with particularly low concentrations of alanine, suggesting that alanine starvation is a likely signal inducing the activation of Gcn2p in this context (Fig. 2-10b). Additionally, methionine and aromatic amino acids showed very low concentration in all strains (Fig. 2-10b), which was in line with their higher biosynthetic costs (47). Amino acid starvation also induces the repression of the target of rapamycin complex 1 (TORC1) (55), which in turn leads to decreased phosphorylation of a large number of proteins (56). In the two high yield strains, there were 14.9% (250 out of 1 676) proteins with decreased phosphorylation events, and given the scale to TORC1 targets, the fraction was 23.6% (13 out of 55) (Fig. 2-10c), indicating that TORC1 was repressed in these strains. Furthermore, we found that all Gcn4p regulated genes (both upregulation and downregulation) were differentially expressed in all protein production strains compared to Qc (Fig. 2-10d), suggesting that Gcn4p was activated in response to protein production.

In addition to Gcn2p-Ser-577, two phosphorylation sites on the complex eukaryotic initiation factor 2B (eIF2B), Gcd1p-Ser-296 and Gcd6p-Ser-449, also showed decreased phosphorylation. Since eIF2B activity is inhibited by phosphorylation of eIF2α and negatively regulates the expression of Gcn4p (Fig. 2-10a), it is likely that the dephosphorylation of these two sites were important for inactivation of eIF2B.

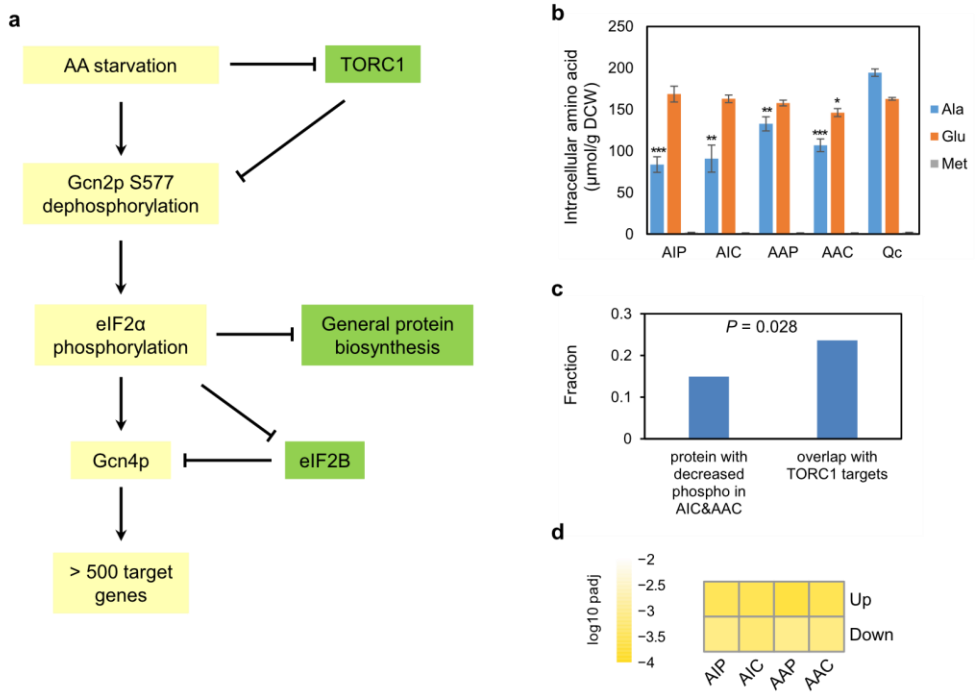

Figure 2-10. Activation of kinase Gcn2p induced the activation of Gcn4p. a The Gcn2p signaling pathway activated by amino acid starvation. **b** Intracellular abundance of alanine, glutamate and methionine in all strains. Data shown are mean values ± standard errors of the means of biological triplicates. Statistical significance was determined by a two-tailed Student's *t* test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. **c** Fraction of dephosphorylated proteins in the high yield strains (AIC and AAC) compared with Qc, and the ratio of the number of these proteins to the number of TORC1-repression induced dephosphorylated proteins. Binominal cumulative distribution function was used to calculate the *P* value. **d** Changes of expression level of gene sets regulated (upregulated or downregulated) by Gcn4p in the protein production strains. Qc was used as the reference.

2.2 The rate-limiting steps inside the secretory pathway

In section 2.1, I show the global reprogramming of central carbon metabolism in response to recombinant protein synthesis and secretion. To further unravel the mechanisms of the complicated protein secretory pathway, recombinant protein with more complex structure and modifications, and yeast strains with higher production yield, are more valuable to be studied. Herein we characterize the strain AAC, the high α-amylase yield strain in section 2.1, and two evolved strains with higher α-amylase yields (Fig. 2-11), to explore the crucial rate-limiting steps for improving recombinant protein production.

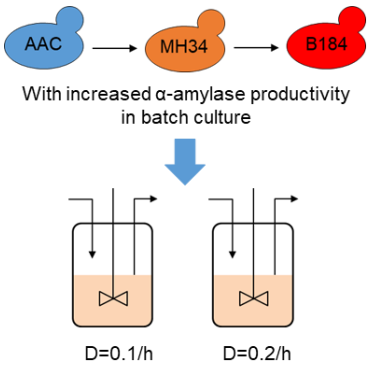

Figure 2-11. Relationships among the α-Amylase production yeast strains used in Paper II. All three strains were grown in chemostat cultures operated at dilution rates of 0.1/h and 0.2/h.

Previous work showed that a higher specific growth rate was coupled to higher recombinant protein production in chemostat cultures (57). In view of the maximum specific growth rates of these three strains (0.276/h, 0.329/h and 0.310/h for AAC, MH34 and B184, respectively) (58), we cultured the strains at dilution rates of 0.1/h and 0.2/h. Compared with the reference strain AAC, we found that the α-amylase productivity of MH34 and B184 was significantly improved at both dilution rates (Fig. 2-12a), which is in agreement with the results obtained for batch cultures (58). However, in batch culture, B184 produced α-amylase with 1.29-fold increase of yield and 47% increase of productivity compared with MH34. In contrast, in chemostat cultures, we found that the yield or the productivity of B184 was only 14% greater than that of MH34 at 0.2/h, and even 22% lower at 0.1/h, indicating the different mechanisms for improved protein production between MH34 and B184. To compare the fraction of misfolded α-amylase between three strains, we also measured the intracellular α-amylase abundance and activity (Fig. 2-12b and c). Compared with B184, we found that in MH34 the abundance was greater but the activity was lower, which means a higher fraction of misfolded α-amylase in MH34 than in B184.

The absolute quantitative proteome for AAC, MH34 and B184 grown in steady-state chemostat cultures at dilution rates 0.1/h and 0.2/h were analyzed. Proteins were annotated based on the Yeast GO-slim bioprocess mapper (48), analyzed the changes of specific bioprocesses in the proteome and listed the processes changing significantly ($P < 0.05$) in MH34 or B184 revealed by proteome analysis (Fig. 2-13). Out of 99 Yeast GO-slim bioprocesses, only 18 processes were differentially expressed, with the scale of $\log_2$ fold change from -0.82 to 1.55. Clearly the proteome changes in a specific manner that precisely impacts bioprocesses related to protein production, including amino acid transport, vitamin metabolism, protein modification, protein phosphorylation and so on.

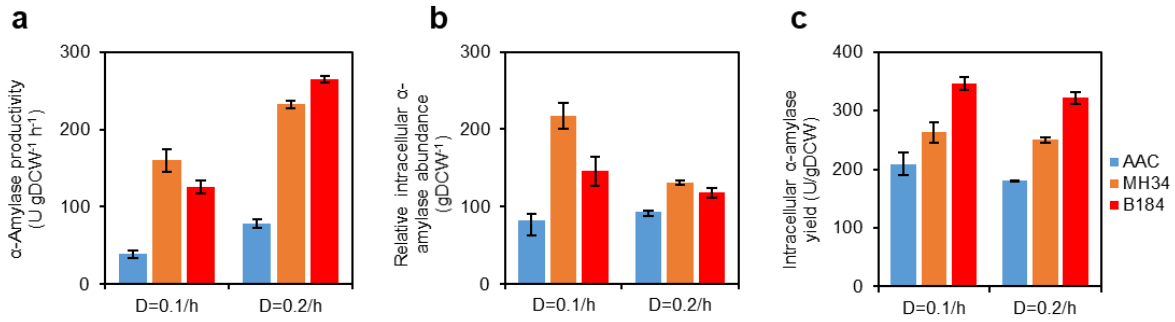

Figure 2-12. α-Amylase production in steady state of chemostat culture. **a** α-Amylase productivity of strains in steady state of chemostat culture. **b** Relative intracellular α-amylase abundance of strains in chemostat culture. **c** Intracellular α-amylase yield of strains in chemostat culture. Data shown are mean values ± standard errors of the means of biological duplicates.

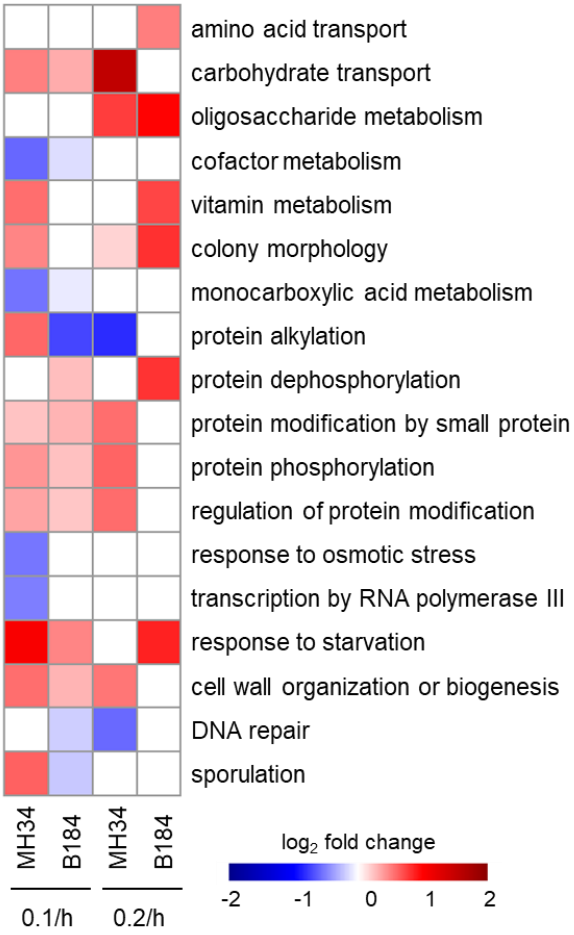

Figure 2-13. The differentially expressed biological processes ($P < 0.05$) in MH34 and B184 revealed by proteome allocation. All proteins were allocated to the 99 Yeast GO-slim biological processes. The proteome fraction (g/g) for each process was used. The expression levels of processes in AAC at 0.1/h and 0.2/h, respectively, were used as the reference.

Since energy metabolism is important for protein synthesis and secretion (59). We next investigated the abundance of proteins involved in carbohydrate transport and oligosaccharide metabolism, which were upregulated in the high α-amylase production strains (Fig. 2-13). The upregulated proteins mainly participate in glycogen metabolism (Fig. 2-14), which links carbohydrate metabolism to protein glycosylation and folding via the production of uridine diphosphate glucose (UDP-glucose) (60). α-Amylase is a glycoprotein which carries one of *N*-glycans. And UDP-glucose is a critical precursor of *N*-glycans, the changes in composition of which act as a signal that guides the folding process for glycoproteins. For example, when the *N*-glycan changes from G3M9 ($Glc_3Man_9GlcNAc_2$) to G1M9 ($Glc_1Man_9GlcNAc_2$) the protein is better folded by relevant enzymes and chaperones, and when it changes from M8 ($Man_8GlcNAc_2$) to M7 ($Man_7GlcNAc_2$) there is increased targeting of the protein to ERAD for degradation (20). Hereby the upregulation in glycogen metabolism is able to offer more available sugars, which could be recruited by *N*-glycans and store more sugars that are trimmed from *N*-glycans, which could support the increased protein production in MH34 and B184.

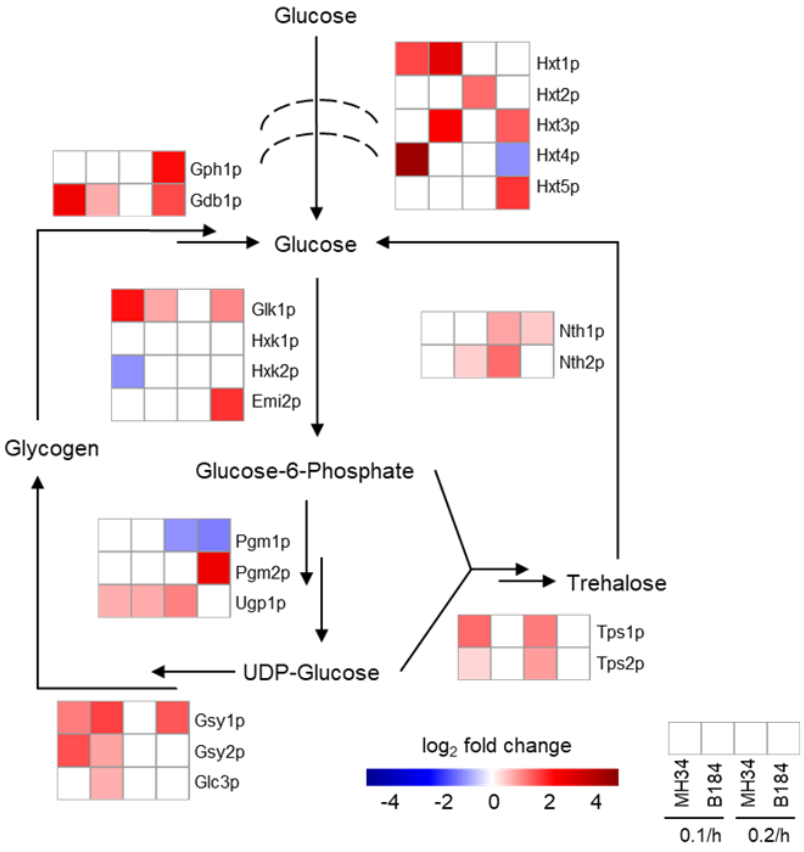

Figure 2-14. Differentially expressed proteins (*P* < 0.05) related to glycogen and trehalose metabolism. The expression levels in AAC at 0.1/h and 0.2/h were used as the reference accordingly.

Differ from glycogen metabolism, trehalose metabolism showed differences between MH34 and B184. The proteins in trehalose metabolism were upregulated mainly in MH34, but not in B184 (Fig. 2-14). Previous studies had shown that trehalose exerts bidirectional effects on protein folding (61, 62). On the one hand, trehalose can help to prevent folded proteins from denaturing and misfolded proteins from aggregating. On the other hand, trehalose interferes with refolding of denatured proteins by relevant molecular chaperones. Therefore, the upregulation of the trehalose cycle in MH34 suggests a dynamic control of the trehalose concentration in the cell, which is line with the fact that there are more misfolded proteins in MH34 than in B184. To test this hypothesis, we calculated the energy used for protein synthesis and secretion based on a genome-scale metabolic model with enzyme constraints (Yeast8) (38). Energy used for protein production is a component of the non-growth associated maintenance energy (NGAM) in Yeast8, which we found to be always lower in B184 than in other strains (Fig. 2-15). Especially compared with MH34, due to their similar α-amylase productivity in chemostat, the lower NGAM in B184 indicated that B184 expended less energy than MH34 to produce the same amount of correctly folded α-amylase, which was in line with the finding that the fraction of misfolded α-amylase was greater in MH34 than in B184.

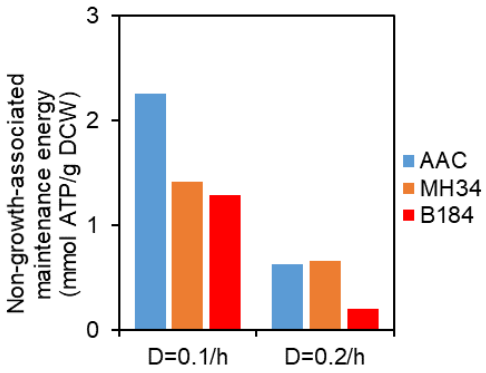

Figure 2-15. The non-growth associated maintenance energy (NGAM) in chemostat cultures as calculated with assistance of the Yeast8 model.

The overall proteome allocation analysis revealed significant differences in protein synthesis associated processes in addition to energy metabolism (Fig. 2-13 and 2-14). We next focused specifically on the protein secretory pathway, which involves more than 160 proteins that are responsible for different post-translational processes in yeast. A previous yeast protein secretory model divided the secretory machinery into 16 subsystems (14). To reduce the complexity, we merged subsystems with similar functions and simplified the secretory pathway into 8 subsystems, including translocation, ER glycosylation, folding, ER

glycosylphosphatidylinositol (GPI) anchoring, ERAD, trafficking between ER and Golgi, Golgi processing and sorting.

Proteome analysis revealed the differences in protein glycosylation and folding processes between strains. As a glycoprotein, the folding pathway is directed by the change of the composition of *N*-glycans (Fig. 2-16). Therefore, we further investigated the involved proteins in these strains to elucidate the specific contributors.

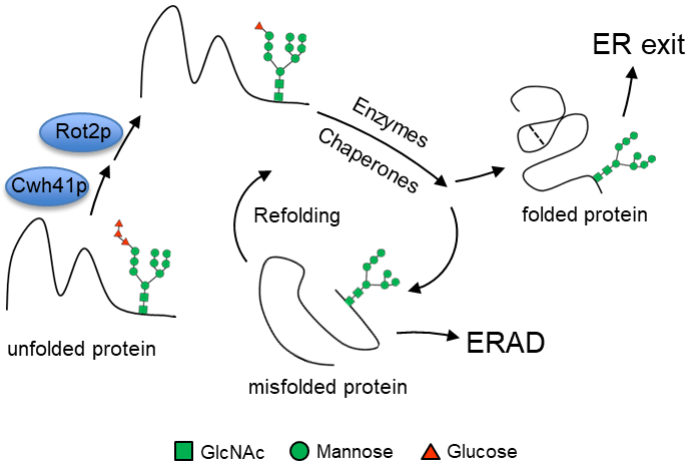

Figure 2-16. *N*-glycan directed protein folding pathway.

Within the ER glycosylation process, there were more downregulated proteins in MH34 than in B184 (Fig. 2-17). Among them, Alg11p and Ost1p, which regulated the synthesis of *N*-glycan (20), are downregulated at both dilution rates in MH34, but not in B184. And correspondingly, within the folding process, the glucosidase Cwh41p, which catalyzes the first step in *N*-glycan trimming and initiates the folding process (Fig. 2-16), was upregulated at both dilution rates in B184, and only at 0.2/h in MH34. On the other hand, Pdi1p was the only protein in the folding pathway to be upregulated at both dilution rates in MH34 (Fig. 2-17). This is in line with its role in disulfide bond formation during protein folding as well as in guiding misfolded proteins to ERAD (20). Previous studies already reported that overexpression of *PDI1* can increase α-amylase production (8, 63).

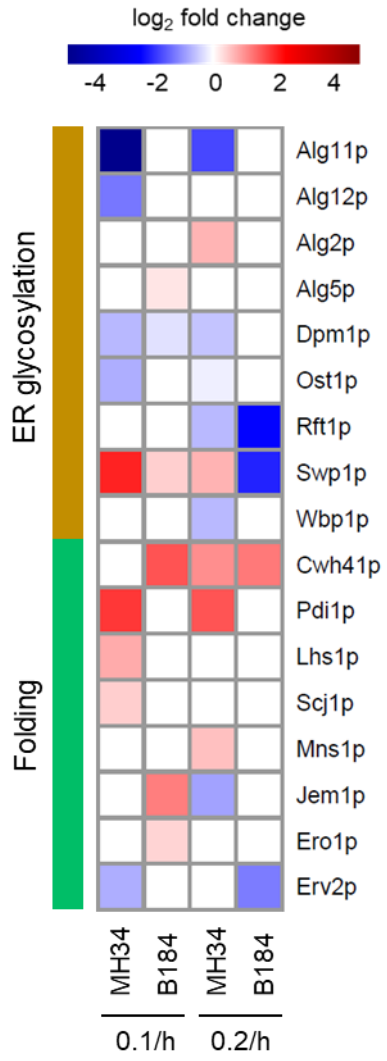

Figure 2-17. Differentially expressed proteins (*P*<0.05) related to the ER glycosylation process and the folding pathway. The expression levels in AAC at 0.1/h and 0.2/h, respectively, were used as the reference.

To validate if the regulation of synthesis and trimming of *N*-glycans could increase the protein yield and cause the differences between MH34 and B184, we studied the role of Cwh41p on protein production. At first, we tested different promoters for overexpression. We overexpressed *CWH41* under the control of either a strong promoter or the *CWH41* native protomer carried by a multi-copy plasmid in AAC. We also tested the overexpression of Rot2p, which catalyzes the second step in *N*-glycan trimming (Fig. 2-16). The α-amylase titer increased with overexpression of *CWH41*, particularly when expressed under its native promoter, reaching a 40% increase in α-amylase production (Fig. 2-18a). Interestingly, the final biomass was significantly increased in this strain as well, which was rarely observed in previous engineering studies (8, 58, 64). In general, the final biomass is most of the times lower in strains engineered to overproduce recombinant proteins, since increased protein

production overloads the folding burden of the cell, leading to accumulation of misfolded proteins and increased energy expenditure in response to cell stress (65). One example is the overexpression of *PDI1*, which increased α-amylase yield but decreased final biomass (8). Here the increase of both α-amylase production and final biomass suggests that *CWH41* overexpression improves protein production without accumulation of misfolded proteins. In other words, *CWH41* overexpression might lead to more precise protein folding, consistent with the decreased fraction of misfolded α-amylase in B184.

A previous study reported that the overexpression of *PDI1* under the control of the *TEF1* promoter resulted in a higher α-amylase yield than overexpression using its native promoter (8). In case of *CWH41* and *ROT2*, the native promoters led to better results (Fig. 2-18a). Therefore, here we performed single-copy combinatorial overexpression in AAC with the superior promoter for each gene to further evaluate the role of Cwh41p (Fig. 2-18b). The data showed that the overexpression of *CWH41* itself did not increase the α-amylase yield in AAC, but the combinational overexpression of *PDI1* and *CWH41* did further increase the yield compared with the overexpression of *PDI1* alone, which could indicate that the increase of folding precision is more important for strains with improved protein folding capacity. To further validate this, we engineered the strain MH34, which has an improved folding capacity compared with AAC. Both of the multi-copy overexpression via plasmid (Fig. 2-18c) and single-copy combinatorial overexpression (Fig. 2-18d) showed that *CWH41* can increase the α-amylase yield. In details, the single-copy combinatorial overexpression revealed that *CWH41* itself can help to increase the α-amylase yield by 27%, compared with the 43% increase in B184 at the same condition (Fig. 2-18d), which is in line with the observation that overexpression of *CWH41* helped to further increase the yield in an AAC strain already overexpressing *PDI1*. For *PDI1*, the expression level was already upregulated in MH34 (Fig. 2-17), and overexpression did therefore not result in a significant difference (Fig. 2-18c and d).

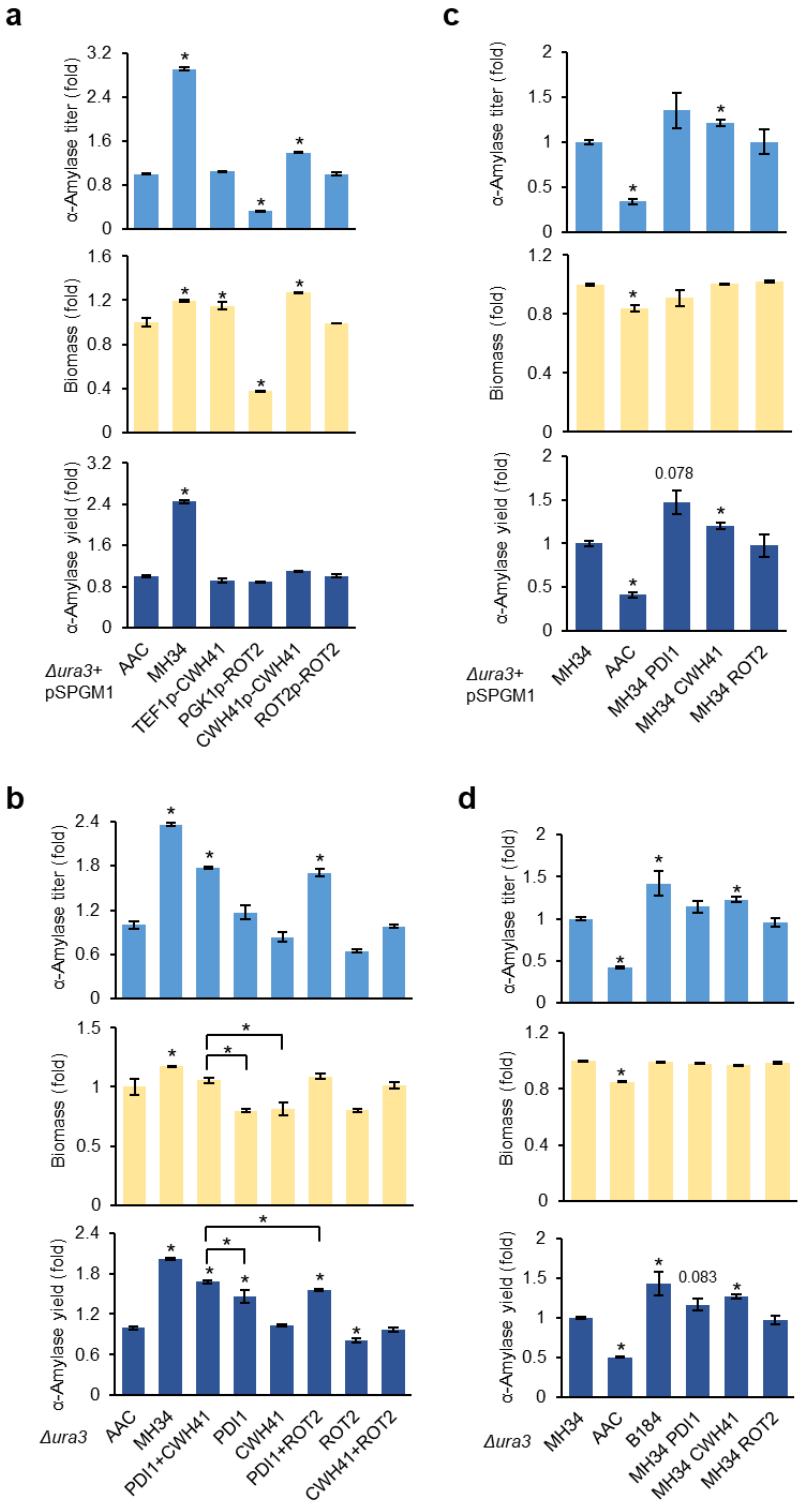

Figure 2-18. Overexpression of *CWH41* contributes to α-Amylase production. a Promoters evaluation for overexpression of *CWH41* or *ROT2* from plasmids. The plasmid pSPGM1 was used for the gene overexpression. **b** α-Amylase titer, biomass and α-amylase yield of engineered strains in the background of MH34. The plasmid pSPGM1 was used for the gene overexpression. **c** α-Amylase titer, biomass and α-amylase yield of engineered strains in the background of AAC. **d** α-Amylase titer, biomass and α-amylase yield of engineered strains in the background of MH34. Data shown are mean values ± standard errors of the means of biological triplicates. Statistical significance was determined by a two-tailed Student's *t* test. *$P < 0.05$.

Proteome analysis suggested that Cwh41p and Pdi1p are two important proteins in this context. Herein we also studied the regulatory mechanisms of these two proteins based on their mRNA abundance, protein abundance and translation propensity (protein abundance / mRNA abundance, abbreviated to P/T) among three strains (Fig. 2-19). For Cwh41p, compared with AAC, at transcriptional level there was limited increase in the evolved strains (less than 5% increase in B184 at 0.1/h, and even decrease under other three conditions), and at the translational level the abundance significantly increased (over 100% increase in B184 at both dilution rates). Thus, the increase of Cwh41p mainly depends on post-transcriptional regulation. And for Pdi1p, there was already significant increase at transcriptional level in the two evolved strains, but the different post-transcriptional regulation led to different protein abundance (over 50% increase of P/T value in MH34 at both dilutions, and decreased P/T value in B184). Thus, the regulation of Pdi1p needs both transcriptional and post-transcriptional regulation.

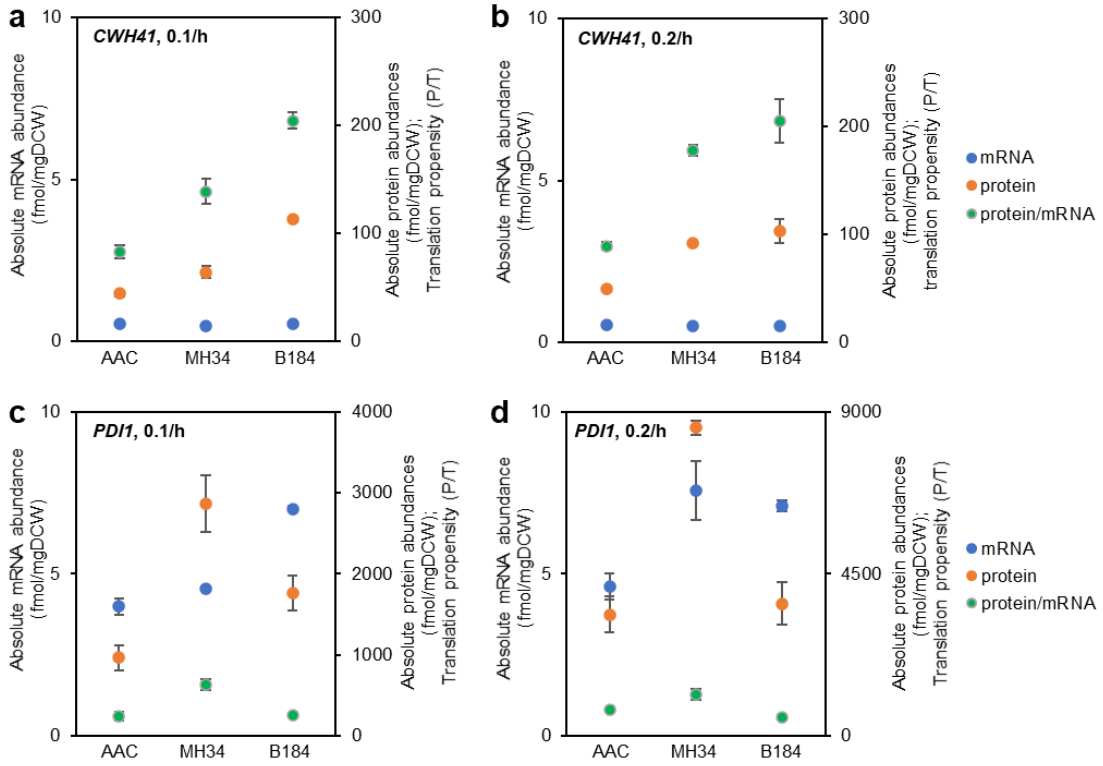

Figure 2-19. The mRNA abundance, protein abundance and translation propensity (protein abundance / mRNA abundance) of *CWH41* or *PDI1* of strains in chemostat cultures. a and **b** show the properties of *CWH41* expression under 0.1/h and 0.2/h, respectively. **c** and **d** show the properties of *PDI1* expression under 0.1/h and 0.2/h, respectively.

3. Translation – the initiation of protein synthesis

The translation process translates mRNA to polypeptide in the organelle ribosome. A quantitative measurement of transcriptome, proteome and abundance of ribosomal proteins enables a more comprehensive understanding of the protein synthesis process. To further characterize the correlation between mRNA and protein, as carbon and nitrogen are two important elements in the composition of nucleotide and amino acid, the supplemented C/N ratio is changed by stepwise reducing the nitrogen content in the growth medium.

The typical elemental composition of yeast dry biomass is ~49% carbon and 9% nitrogen by weight (66), which represents a C/N ratio of 5.4. In typical carbon-limited chemostat cultures, nitrogen is provided in excess with C/N ratio of 3-4 in the growth medium (67, 68). Reducing the nitrogen content in the growth medium to a C/N ratio of 50-115 leads to total protein and mRNA in the cell declining to 50% and 25% of the amount measured in C-limited cultures, respectively (Fig. 3-1). Further reducing nitrogen availability reduced the steady-state biomass, but did not further decrease the RNA and protein fraction, suggesting these levels to be the minimum required for cell growth at a constant dilution rate of 0.2/h (Fig. 3-1). Under these nitrogen-limited conditions therefore, we consider the transcriptome and proteome allocation of the cell to be fully economized.

Cells maintain reserves in translational capacity, including but not limited to excess/inactive ribosomes (69–71). In our dataset, since the total protein-to-mRNA ratio was doubled under N-limitation (Fig. 3-1), this places the overall translational reserve at approximately 50% capacity in C-limited cultures; a conservative estimate, as this approximation neglects protein turnover. As both the transcriptome and proteome allocation for different cellular processes maintained a similar % of reserves, we next investigated whether this is true for cellular reserves in translational capacity as well.

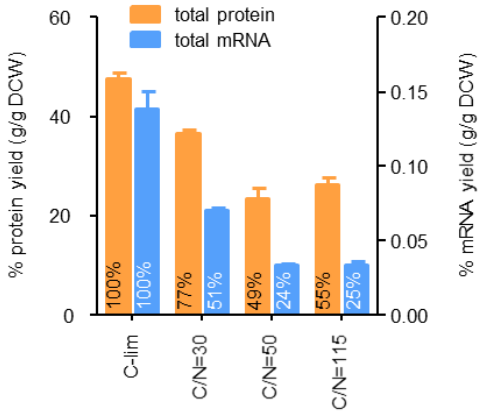

Figure 3-1. Total protein and mRNA content of the biomass under each culture condition. Bars are mean + standard deviation of biological duplicates.

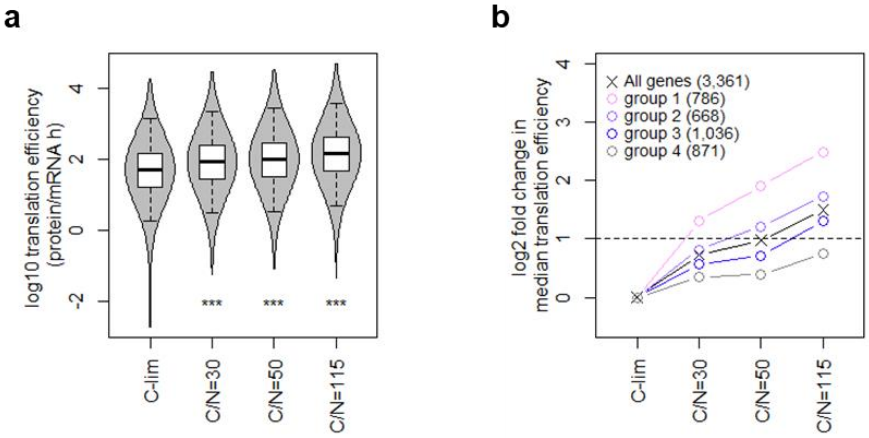

Figure 3-2. Reserves in translational capacity. a Gene-specific translation efficiency was calculated in all growth conditions. ***, $P < 0.001$, Student's *t* test. **b** Genes were grouped based on the step of nitrogen reduction at which their translation efficiency was increased by log2 > 1. Data points reflect the median translation efficiency of each group, with the # of genes belonging to each group shown in parentheses. Horizontal dashed line indicates 2-fold increase from C-limited cultures.

Using protein and mRNA abundances from this dataset, and by mining protein turnover data from Lahtvee *et al* (2017) (68), we calculated protein synthesis efficiency k_{sP} (protein/mRNA h) for each protein. With step-wise reduction of nitrogen content in the growth media, k_{sP} was globally increased (Fig. 3-2a), indicating that reserve translational capacities were placed into usage: a 2-fold increase in the k_{sP} of a gene implicates a 50% reserve translational capacity for this gene under typical C-limited growth. To study this response, we further grouped genes based on the step of nitrogen reduction at which k_{sP} was increased by >2-fold, which

splits a total of 3,361 genes into 4 groups of similar sizes (Fig. 3-2b). The step of k_{sP} modulation correlated with the effect size: those genes with an increased k_{sP} at C/N=30 (Fig. 3-2b, group 1), were also the genes with the largest k_{sP} increase at C/N=115. When nitrogen availability is at its lowest, at C/N=115, this group of genes has a median k_{sP} increase of 5.6-fold (Fig. 3-2b), implicating an 82% reserve translational capacity under typical C-limited growth. Remarkably, this group of genes were enriched exclusively in processes related to metabolism (Fig. 3-3, group 1). This is consistent with our observation that the abundance of enzymes within a given metabolic pathway have a higher propensity to be internally adjusted, while proteome and transcriptome allocation to the pathway as a whole remain constant. Taken together, these results show that the vast majority of cellular reserves are dedicated to metabolism, in line with the growing consensus that yeast have evolved complex regulatory mechanisms to control metabolic pathways.

At the final step of nitrogen reduction in the growth medium, a total of 2,490/3,361 genes (74%) exhibited a >2-fold k_{sP} increase, demonstrating >50% reserve in their translational capacities when growing in C-limited conditions (Fig. 3-2b). These genes were enriched for a variety of cellular processes (Fig. 3-3). The remaining 26% of genes showed little to no modulation of k_{sP} and were enriched in translation/protein processing-related GO-slim terms (Fig. 3-3). This indicates that components of the translational machinery were themselves being translated at maximum capacity in C-limited cultures, and therefore the changes in the abundance of protein related to these processes are predominantly regulated at the transcript level. Finally, we observed no significant enrichment ($p_{\text{Fisher}} > 0.01$) in any of the 4 groups herein for essential genes (48), indicating that the unequal distribution of these reserves is not a critical feature for survival, but likely arose by conveying a selective advantage and increased fitness.

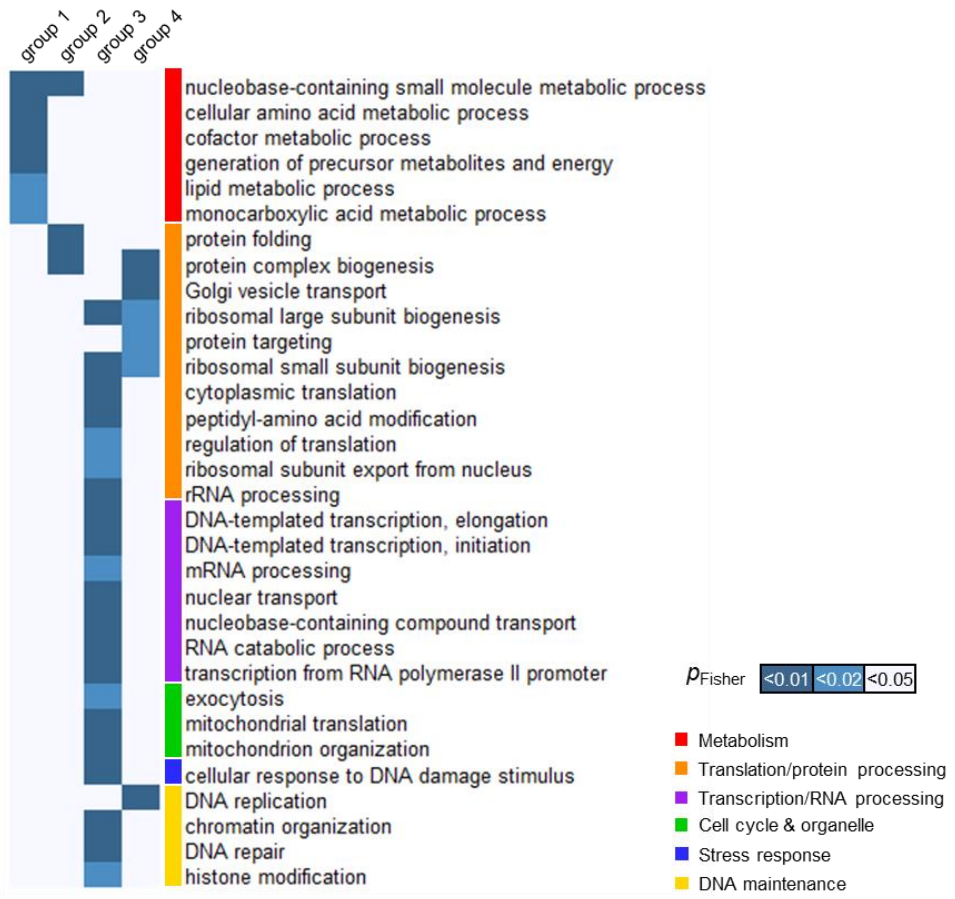

Figure 3-3. Reserves of translational capacity are preferentially used to translate metabolic proteins. Enrichment of GO-slim terms in each gene group as defined in Fig. 3-2b was analyzed by Fisher's exact test. GO-slim terms with FDR-adjusted $p_{\text{Fisher}} < 0.05$ in at least one gene group are shown.

The observation that cells preferentially maintain reserve translational capacities for some processes (metabolism) but not others (components of the translational machinery) is interesting, considering that functionally distinct sub-pools of mRNA are known to be translated by subsets of ribosomes with distinct ribosomal protein (RP) stoichiometry (72–74). RP stoichiometry has also been shown to depend on the balance between the economics of protein production and cell growth (75). To investigate this further, we examined RP stoichiometry in our proteomics dataset for C-limited cultures, and found that of the 76/79 RP subunits detected (from 94/117 non-identical RPs), their abundance spanned >2 orders of magnitude, even when their paralogs were summed. This large variation could not be accounted for by differences in RP degradation rate (68), molecular weight, order of assembly (76); or parameters of the MS such as the number of peptides detected or the coverage of each RP's protein sequence. Interestingly, most deviations from mean RP abundance and rRNA abundance were sub-stoichiometric RPs, while few RPs were over-represented (Fig.

3-4a). This indicates that a substantial number of ribosomes in the cell likely contain a sub-stoichiometric composition of RPs that are available.

To validate this finding, we performed targeted quantitative proteomics for 26 RPs by TMT-based MS, using 49 proteotypic peptides that were chemically synthesized at known quantities as standards. After tryptic digest and labeling with isotopic mass tags, the synthetic peptide standards and the pooled reference sample were multiplexed at 8 different ratios to ensure that MS intensity ratios cover the dynamic range of TMT (77). This analysis confirmed that iBAQ-based quantification is robust to the order of magnitude, for 23 out of 26 (88%) RPs. We note here that, for a few RPs, the measured abundance of different peptides differed by >10-fold, e.g. for RPL30 and RPL37A, which could have arisen from partial degradation during sample preparation and handling. In all cases, the average of all peptides was taken as the abundance of the RP. This analysis further indicated that any post-translational modifications on these peptides did not significantly interfere with the absolute quantification by MS, since the synthetic peptide standards are not modified. Thus, we confirmed that the abundance of RP subunits in a cell is markedly different from the typically assumed 1:1 stoichiometry. Out of the 76 RP subunits detected in our dataset, 54 (70%) were expressed in the order of 10^5 molecules/cell (mean = 3 x 10^5 molecules/cell), in line with classic estimates of ribosome content (78); while 22 (30%) RP subunits were sub-stoichiometric by up to an order of magnitude.

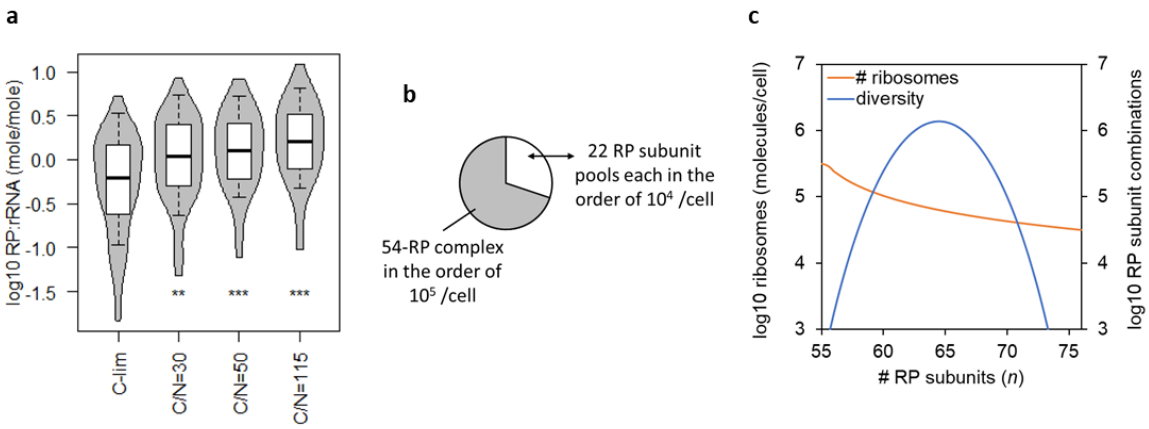

Figure 3-4. Ribosome reserves contain diverse sub-stoichiometric ribosomal proteins (RP). a RP:rRNA ratio for each RP was calculated in all growth conditions. **, $P < 0.01$; ***, $P < 0.001$, Student's *t* tests. **b** Model of ribosome complex diversity and abundance. Each complex ribosome contains 54 "core" RP subunits expressed in the order of 10^5 molecules/cell, and samples a subset of 22 RP subunits expressed at 10^4 molecules/cell. **c** The total number of ribosomes in the cell decreases, while diversity of ribosomes varies parabolically, with the number of RP subunits complexed.

From this data, a simple model of ribosome complex diversity and abundance emerges (Fig. 3-4b and c). Each complexed ribosome likely contains the 54 "core" RP subunits that are expressed in the order of 10^5 molecules/cell, and a subset of the 22 RP subunits, expressed in the order of 10^4 molecules/cell (Fig. 3-4b). Thus, the total number of ribosomes that can be complexed with *n* RP subunits decreases with increasing *n* when $n > 54$ (Fig. 3-4c). The diversity of ribosome subunit composition varies parabolically with *n*, with peak diversity occurring with complexes of n = 64-65 RP subunits, where $>10^6$ combinations are theoretically possible (Fig. 3-4c). Notably, for complexes of $n = 58$ to $n = 71$, the number of possible RP subunit combinations exceeds the actual number of ribosomes that can be built with *n* RP subunits (Fig. 3-4c). This means that it is possible for these ribosomes to re-assemble into completely new RP subunit compositions if needed, pointing to a dynamic pool of ribosome reserves that can quickly respond to changing growth environments by adjusting their RP subunit stoichiometries.

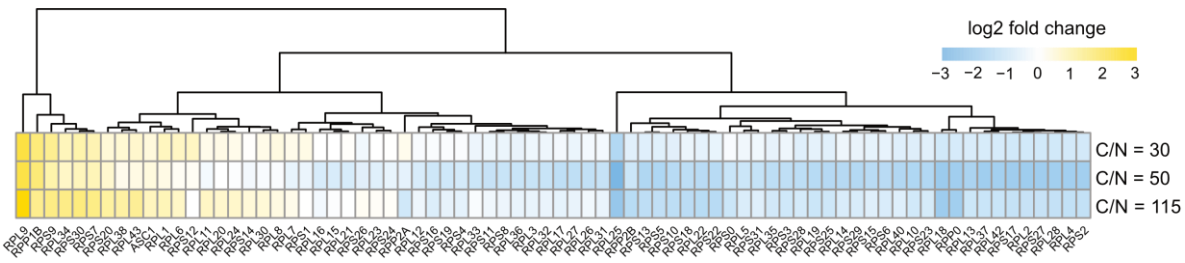

Figure 3-5. Relative abundance of RP subunits in N-limited chemostats compared to C-limited chemostat. RP paralogs were summed and labeled with standard gene name.

As nitrogen is reduced in the growth media, less RPs are detected at sub-stoichiometric levels to rRNA (Fig. 3-4a), with 17 RP subunits (22 RPs) being selectively upregulated by >2-fold (Fig. 3-5). This upregulated RP pool includes 3 RP subunits where one paralog of the paralog-pair (RPL21B, RPL6A, and RPS7A) was upregulated, while the other (RPL21A, RPL6B, RPS7B) was not, indicative of a paralog-specific response to nitrogen reduction. Interestingly, not all RPs upregulated in the N-limited cultures were sub-stoichiometric in the control C-limited culture, and vice versa, implicating multiple levels of control for RP subunit incorporation and ribosomal translation efficiency. Benchmarking the quality of this gene set is the inclusion of Asc1 (Fig. 3-5), a peripheral RP in the small subunit that is known to promote ribosomal activity. In particular, such activity is targeted towards highly translated mRNAs containing short ORFs (79), with Asc1p interacting with the pre-initiation complex

(80), the closed-loop complex (79), and RNA-binding proteins (RBPs), to tether mRNAs to the translational machinery (81, 82). These adjustments in ribosome composition (Fig. 3-5) may therefore be responsible, at least in part, for the global increase in k_{sp} under nitrogen limitation (Fig. 3-2a), by helping sub-stoichiometric ribosome reserves engage with and translate mRNA, with a preference for those encoding metabolic enzymes (Fig. 3-3). Taken together, these observations suggest the intriguing possibility that RP stoichiometry can be optimized to improve translation efficiency for synthetic biology and metabolic engineering applications.

4. Proteome-constrained secretory modeling

The latest yeast genome-scale metabolic model (GEM) Yeast8 (38) represents a comprehensive computational resource for performing simulations of the metabolism in yeast, the applications of which has been shown in Chapters 2 and 3. To further elucidate the secretory mechanisms and predict genomic targets for improving recombinant protein production, herein I introduce the construction of a proteome-constrained genome-scale secretory model (pcSecYeast). In pcSecYeast, the protein secretory pathway was elaborated by adding a series of relevant metabolic reactions, including reactions for the generation of precursors from the central carbon metabolism, e.g. GPI and glycans, and the central reactions within the secretory pathway, e.g. protein translocation, glycosylation, folding, ERAD and sorting (Fig. 4-1), conceptually similar with the earlier published metabolic models for *E. coli* (83) and *S. cerevisiae* (84). Therefore, the model can simulate the whole process, including nascent polypeptide formation, protein maturation and protein transport to various destinations. To our knowledge, it is the first genome-scale metabolic model to comprehensively describe the protein synthesis and secretion process in yeast. Overall, pcSecYeast accounts for 1, 639 protein-coding genes and approximately 70% of the total proteome mass according to paxDb.

As ribosome and protein synthesis demands substrates and energy from cellular metabolism, and the synthesized enzyme complexes are crucial for the catalysis of metabolic reactions, herein we introduced coupled constraints to correlate protein synthesis and cellular metabolism. Additionally, the synthesis of enzyme complexes also constrains the protein synthesis and secretion process, including ribosome synthesis and the synthesis of components involved in the secretory pathway. Moreover, in pcSecYeast, the flux of enzymatic reaction is constrained by the turnover number and the concentration of the enzyme, allowing the simulation of the minimum protein demand that maintains the metabolic status, i.e., the proteome-constrained metabolic status.

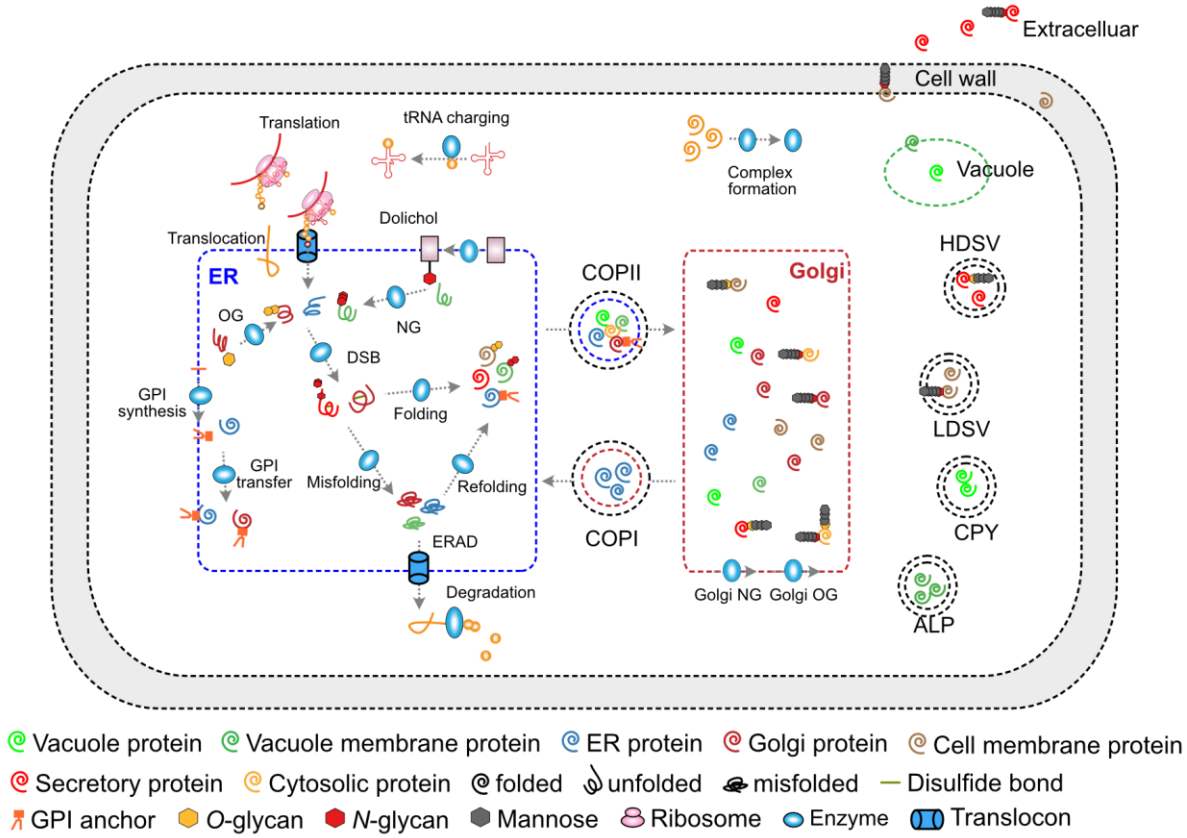

Figure 4-1. Simplified schematic processes involved in the secretory pathway. The pathway includes protein translation, translocation, glycosylation, GPI transfer, ERAD and sorting processes. NG: *N*-glycosylation, OG: *O*-glycosylation, DSB: disulfide bond formation, GPI: glycosylphosphatidylinositol, ERAD: ER-associated degradation, LDSV: low-density secretory vesicles, HDSV: high-density secretory vesicles, ALP: alkaline phosphatase pathway, CPY: carboxypeptidase Y pathway.

As mentioned previously, high-fidelity protein folding is a crucial pathway in the secretory machinery. Misfolded proteins, which can be caused by many internal or external factors, need to be refolded or delivered to the proteasome for degradation through ERAD pathway (20). Without proper handling, the accumulation of misfolded proteins can trigger cell disorder (85–87). Here, we simulated the process of vacuolar carboxypeptidase Y (CPY) production (88) to characterize the ER tolerance to the accumulation of misfolded proteins. In the model, the misfolding level of CPY is simulated by modifying the misfolding ratio. A misfolding ratio of 1 means that all CPY molecules are fully misfolded and cannot be targeted to the Golgi for further processing, as reported by Haynes *et al.* (89). The maximum growth reduction was used to characterize the cellular fitness cost when expressing CPY with different misfolding ratio. We found that there was ~1.9% maximum growth reduction when expressing 0.46 mg native CPY without misfolding (representing 0.1% of the total proteome, Fig. 4-2), which is comparable to a previous study that simulates the expression of the yellow

fluorescent protein (YFP) (90). The simulation showed that a higher misfolding ratio or a longer ER retention time led to a higher maximum growth reduction, i.e., a higher fitness cost.

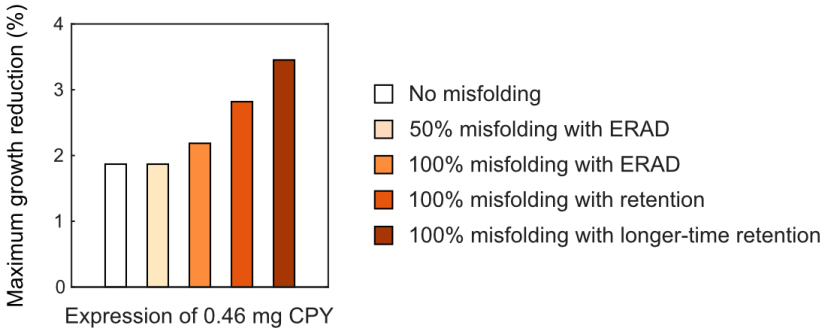

Figure 4-2. Reduction of simulated maximum specific growth rate (1/h) due to expression at certain level of CPY following different routes.

Furthermore, we also investigated the correlations between CPY expression, misfolding and cell growth (Fig. 4-3). It is shown that when CPY was expressed at high levels, the high misfolding ratio would cause a steep decrease of cell growth (Fig. 4-3a). In addition, the misfolded proteins were prior to be degraded rather than accumulated (Fig. 4-3b and c). However, when the amount of misfolded proteins exceeded the degradation capacity, a steep increase of accumulation rate was observed (Fig. 4-3c).

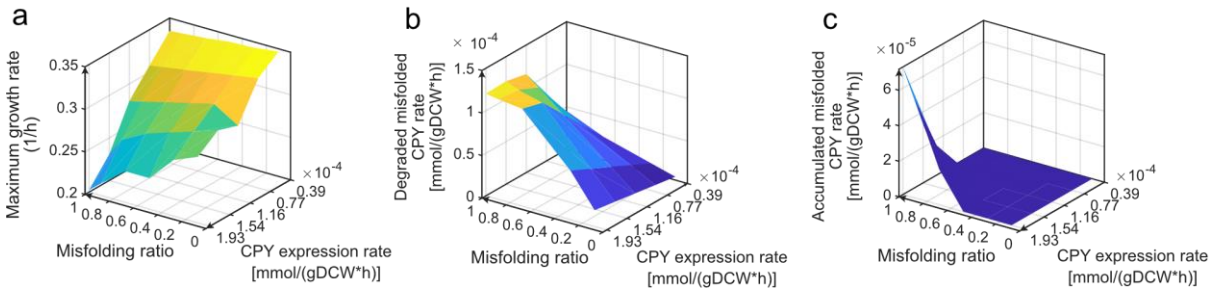

Figure 4-3. Simulation of physiological rates with varying CPY expression level and misfolding ratio.

In addition to the characterization of the protein synthesis and secretion process, the model also can be used to identify the rate-limiting steps for improving recombinant protein production. Here, genomic target prediction was performed through flux scanning based on enforced objective function (FSEOF) (91), where the model was constrained with a stepwise decrease in the specific growth rate, and recombinant protein production was maximized (Fig. 4-4a).

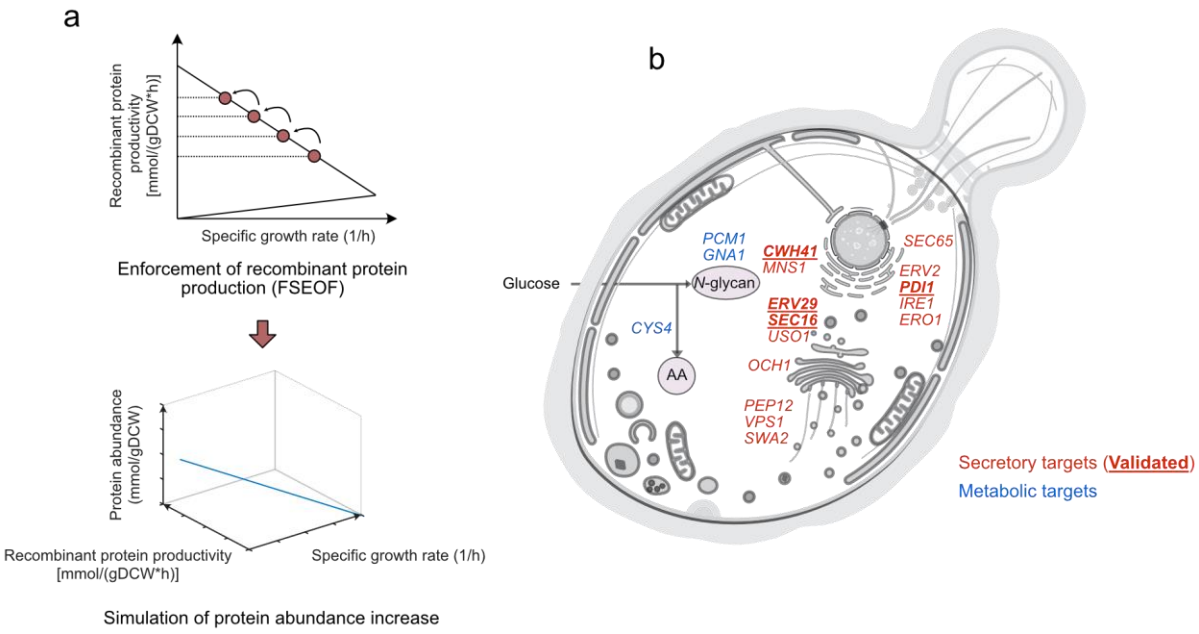

Figure 4-4. Prediction of overexpression targets for improving α-amylase production. a Adapted FSEOF method used for target identification. FSEOF: Flux Scanning based on Enforced Objective Function. **b** Protein localization of the predicted overexpression targets. Yeast compartmentalized figure source: SwissBioPics.

We used α-amylase as model recombinant protein and performed the target prediction. The predicted targets were divided into different groups based on their functions relevant to protein synthesis and secretion, and in each group the most promising targets were chosen based on their predicted values. In total, 17 targets were selected, including 14 secretory targets and 3 metabolic targets. The 14 secretory targets are involved in protein translocation, protein folding and sorting, and the 3 metabolic targets are related to *N*-glycan synthesis and amino acid synthesis (Fig. 4-4b). Next, we performed single-gene overexpression to validate if these targets are important for improving α-amylase production. Among these targets, overexpression of *CWH41* (92), *ERV29* (8), *SEC16* (64) and *PDI1* (8, 93) has already been reported to be useful for improving α-amylase production. The effects of remaining 13 targets were shown in Fig. 4-5.

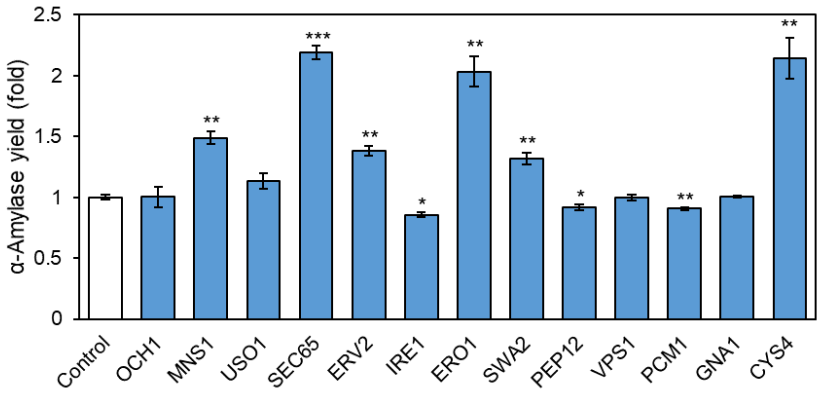

Figure 4-5. Validation of overexpression targets for improving α-amylase production. Data shown are mean values ± standard errors of the means of biological triplicates. Statistical significance was determined by a two-tailed Student's *t* test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

It was indicated that the overexpression of *MNS1*, *SEC65*, *ERV2*, *ERO1*, *SWA2* and *CYS4* significantly increased the α-amylase yield, separately. Among them, Mns1p is responsible for the removal of one mannose residue from a glycosylated protein. The multiple *N*-glycosylation sites of α-amylase suggested the important role of Mns1p on its production. Sec65p is one out of six subunits of the signal recognition particle (SRP), which is involved in protein targeting to the ER (94). Overexpression of Sec65p is anticipated to increase the SRP dependent co-translational translocation, which would benefit α-amylase translocation from cytosol to the ER. Erv2p and Ero1p are involved in the protein folding process in the ER (65). Additionally, overexpression of Ero1p is proved to enhance disulfide-bonded human serum albumin (HSA) secretion in *Kluyveromyces lactis* (95) and single-chain T-cell receptors (scTCR) and single-chain antibodies (scFv) secretion in *S. cerevisiae* (96), suggesting it as a generic target for secretory protein production. Swa2p is involved in vesicular sorting, which is an important step for protein secretion. Cys4p is involved in cysteine synthesis. In the amino acid composition of α-amylase, the fraction of cysteine is 9-fold higher than the average cysteine demand of yeast proteome (58), which demonstrates the importance of cysteine synthesis in α-amylase production. In total, for the identified targets in the secretory pathway, the accuracy was 9/14, while for the 3 metabolic targets, the accuracy was 1/3. The higher proportion and higher accuracy of the secretory targets probably indicates that the rate-limiting steps for recombinant protein production are more related to the secretory pathway rather than other metabolic pathways.

5. Conclusion

Protein synthesis and secretion is a vital process to maintain cell function. Here, we systematically characterize the protein synthesis and secretion process in *Saccharomyces cerevisiae* through integrative analysis of multi-omics data and genome-scale metabolic modeling. In **Paper I** and **Paper II**, through characterization of the recombinant protein production strains, we quantified the cellular resource reallocation in response to recombinant protein production and identified the underlying mechanisms in the protein secretory pathway for improving recombinant protein production, respectively. Surrounding the protein synthesis process, we also quantified the reserves in translational capacity and found that a major part of the translation capacity reserves is preferentially maintained for metabolic processes (**Paper III**). Moreover, we integrated the protein translation, protein post-translational modifications, ERAD and sorting processes into a genome-scale metabolic model of yeast (Yeast8) to construct the pcSecYeast model, enabling simulation of the protein synthesis and secretion process and prediction of high-accuracy genomic targets for improving recombinant protein production (**Paper IV**). Our study also demonstrates the combination of multi-omics data analyses and metabolic modeling to dissect complex cellular responses and intertwined regulatory systems.

6. Future perspectives

6.1 Cofactor supply in protein synthesis and secretion

The yeast *S. cerevisiae* is a popular cell factory that has been engineered to produce many kinds of chemicals and recombinant proteins. The central carbon metabolism is usually reshaped to meet the production demand of various chemicals (5, 97), herein we found that in response to improving recombinant protein production, cell metabolism was reprogrammed to provide more cytosolic NADPH and amino acids (**Paper I**). Energy and redox status, which are typically represented by relative levels of cofactor pairs such as ATP/ADP, NADPH/$NADP^+$ and NADH/NAD^+, are two key elements leading to the remodeling of metabolism (4). In energy metabolism, there is a trade-off between respiration and fermentation to balance the ATP yield and protein cost in response to dynamic environments (98). In this study, the decreased fermentation and the uncoupling between TCA cycle and oxidative phosphorylation indicated that the supply of cytosolic redox capacity prior to energy was critical for the metabolic reprogramming for protein production, which is in line with the severe oxidative stress caused by recombinant protein production (65). Therefore, the accumulation of glycerol in the protein production strains mainly contributed to the maintenance of redox potential, which is also found in previous studies (64, 99). Interestingly, we did not observe any increase in the fraction of proteome allocated to the PPP, which is typically considered to be a crucial pathway for cytosolic NADPH supply. The overall decreased abundance of mitochondrial components potentially indicates a presumable decrease in the mitochondrial volume, which is also observed in a previous study (64). The oxidative phosphorylation process consumes redox power to generate energy and increases cellular oxidative stress, and this could lead to decreased mitochondrial components, considering the high oxidative stress in protein production strains (7, 65). Meanwhile, the increased enzyme usage on TCA cycle and amino acid biosynthesis showed metabolic homeostasis and revealed the reserved capacities maintained by metabolic pathways.

6.2 The regulatory network

Cells have evolved global regulatory network to ensure that metabolic homeostasis is maintained. Learning how the regulation works enables a deep understanding of the cell systems and provides us novel insights which can be leveraged in synthetic biology and metabolic engineering studies. Phosphorylation is an important post-translational modification that is widely involved in signal transduction and regulation (100). In eukaryal cells, the two most important regulators of metabolism are AMP-activated kinase (AMPK) and target of rapamycin complex 1 (TORC1) (101), which regulate metabolism through phosphorylation of enzymes. In **Paper I**, by examining the phosphoproteomic abundance, we found that processes related to translation were highly regulated by phosphorylation in response to increased protein production. Further investigation showed that the amino acid starvation-induced Gcn2p-mediated signaling pathway is a central link between protein production and central carbon metabolism, which demonstrates the importance of phosphoproteome analysis since the abundance of the proteins involved in this pathway kept constant. Additionally, many other types of post-translational modifications, e.g. acylation and sumoylation, can alter protein function as well. Therefore, the combination of protein post-translational modification information with traditional omics data, e.g. proteome and transcriptome, can provide a more comprehensive understanding of cell status, which also calls for the development of high-throughput techniques for quantitative measurement of protein post-translational modifications.

6.3 Novel insights into the secretory pathway

Characterization of cell factories for recombinant protein production is a feasible approach for unraveling of the underlying mechanisms in the secretory pathway. Through modulating the recombinant protein yield (*from zero to one* and *from one to hundreds*), we successfully identified the metabolic reprogramming on the central carbon metabolism in response to recombinant protein production (**Paper I**) and the important role of post-translational modifications on protein production, especially the folding quality control system (**Paper II**). The combined results show that it is necessary to consider both the central metabolism, i.e. the supply of redox power, amino acids and energy, and the post-translational modifications, for improving recombinant protein production. In fact, there is also a link between these two strategies, namely balancing of redox power. As the oxidative stress caused by recombinant protein production is mainly due to the protein folding process and the degradation of misfolded proteins (65), the increased cytosolic NADPH supplied by TCA cycle also

highlights the impact of post-translational modifications on metabolic reprogramming. To date, efforts for improving protein production have primarily targeted the components of the secretory pathway, i.e. proteins involved in the folding pathway or the trafficking process (8, 102–104). Our studies showed that the specific cofactor generation pathway and the specific signaling pathway are also promising alternatives for improving protein production.

Protein folding process is often regarded as the rate-limiting step in the secretory pathway. **Paper II** indicates that, in addition to protein folding capacity, protein folding precision also can be modulated by cell engineering, and its improvement can lead to higher recombinant protein yield. As there are many human diseases caused by protein misfolding (23), including Alzheimer's disease, Parkinson's disease and type II diabetes. Further improving the protein folding precision may benefit the development of new drugs that are used for treatment of these protein misfolding diseases.

In addition, the model pcSecYeast provides a novel platform for the elucidation of underlying mechanisms related to the secretory pathway (**Paper IV**). The simulation of the cellular response to protein misfolding revealed the causation of ER stress. At present, the model incorporates most fundamental processes in the secretory pathway, including protein translocation, *N*-glycosylation, protein folding and ER-associated degradation, further extension of the model to incorporate the unfolded protein response (UPS) and the Golgi processing could potentially increase the simulation accuracy.

6.4 Protein translational reserves

There is growing evidence in both *E. coli* (69, 105) and yeast (71) that cells maintain reserve capacities in protein translation process, trading off maximum exponential growth rate for the ability to respond quickly to changes in their growth environment. In **Paper III** we quantified for the first time the sizes of reserves of the transcriptome, proteome, translational capacity, without the confounding factor of a changing growth rate. Exposing cells to limited nitrogen resources revealed that yeast growing in typical laboratory conditions maintain an overall 50% reserve proteome, 75% reserve transcriptome, and 50% reserve translational capacity. Our analysis showed that a major part of these reserves is preferentially maintained for metabolic processes, highlighting the importance of a robust metabolism for cell growth and survival. Additionally, our data revealed that a global change in k_{sP} is accommodated by adjusting the ribosomal protein stoichiometry, supporting the idea that a large diversity of ribosomes could be present in the cell and be responsible for large differences in gene-specific

k_{sP} (82, 106). Indeed, several landmark studies have shown that ribosomes with or without a specific RP can translate functionally distinct sub-pools of mRNA (72–74), collectively known as the ribosome code (107, 108). We show here with high confidence that 17 RP subunits (22 RPs) are selectively upregulated under nitrogen-limited conditions and are associated with preferential translation for metabolic genes, expanding the ribosome code as well as our understanding of the protein translation process.

Acknowledgements

First and foremost, I would like to thank my supervisor, Jens, for providing me with the opportunity to be part of the SysBio family, for always supporting me with fantastic ideas and suggestions, for encouraging me throughout my studies. You have been a great mentor to me and I feel so privileged to have worked with you.

Thanks to my two former co-supervisors Martin and Dina for your many helpful suggestions and comments on my research projects. Thanks to my co-supervisor Ievgeniia for your many helpful suggestions on my PhD studies and manuscript preparations.

During my studies, I have had the valuable opportunities to work with many great people. Thanks to Rosemary for many helpful discussions and thorough review of my manuscripts. Thanks to Feiran for incredible modeling work in many of our collaborations. Thanks to Kate for your kind and patient helps and discussions when I started my first project. Thanks to Egor and Johannes at GU for the collaborations on proteome analysis. Thanks to John for allowing me to join your amazing project. Thanks to Verena, Ed, Hongzhong, Yuping, Lucy and Veronica for your valuable discussions in countless subgroup meetings. Thanks to Yun, Tao Y. and Fredrik for your valuable helps during Dasgip fermentation. Thanks to Quanli, Xiaowei, Yanyan, Jiwei, Yating, Yi, Jichen, Yu, Xin, Peishun, Tyler, Johan, Michi, Zhiwei, Guokun and Zongjie for many helpful discussions. Thanks to research engineers and administrators, Angelica, Emelie, Marie, Yili, Abder, Elin, Anne-Lise, Martina, Gunilla, Erica, Josefine, Zerif and Anna for your professional assistance in the lab and office.

I would also like to thank many colleagues and friends for their helps during these four years. Thanks to Ivan, Christer, Joakim, Boyang, Lei, Li, Ning, Hao W., Hao L., Demi, Jing, Le, Gang, Rui, Olena, Yassi, Max, Dany, Oliver, Marta, Carl, Francesca, Aleksej, Shaq, Mihail, Florian, Rasool, Filip, Santosh, Mikael, Naghmeh, Xiang, Shadi, Iván, Ela, Linnea, Andrea, Angelo, Jun, Louis, Zhengming, Chunjun, Tao W. and Lingqun.

Last but not least, I owe my deepest gratitude to my family and my parents. Thank you for your unconditional love and support. I love you.

References

1. J. Steensels, K. J. Verstrepen, Taming wild yeast: Potential of conventional and nonconventional yeasts in industrial fermentations. *Annu. Rev. Microbiol.* **68**, 61–80 (2014).

2. S. Marsit, S. Dequin, Diversity and adaptive evolution of Saccharomyces wine yeast: a review. *FEMS Yeast Res.* **15**, 1–12 (2015).

3. J. Peter, *et al.*, Genome evolution across 1,011 Saccharomyces cerevisiae isolates Species-wide genetic and phenotypic diversity. *Nature* **556**, 339–347 (2018).

4. J. Nielsen, Systems Biology of Metabolism. *Annu. Rev. Biochem.* **86**, 245–275 (2017).

5. J. Nielsen, J. D. Keasling, Engineering Cellular Metabolism. *Cell* **164**, 1185–1197 (2016).

6. M. Li, I. Borodina, Application of synthetic biology for production of chemicals in yeast Saccharomyces cerevisiae. *FEMS Yeast Res.* **15**, 1–12 (2015).

7. J. Hou, K. E. J. Tyo, Z. Liu, D. Petranovic, J. Nielsen, Metabolic engineering of recombinant protein secretion by Saccharomyces cerevisiae. *FEMS Yeast Res.* **12**, 491–510 (2012).

8. M. Huang, G. Wang, J. Qin, D. Petranovic, J. Nielsen, Engineering the protein secretory pathway of Saccharomyces cerevisiae enables improved protein production. *Proc. Natl. Acad. Sci.* **115**, E11025–E11032 (2018).

9. G. Palade, Intracellular Aspects of the Process of Protein Synthesis. *Science (80-.).* **189**, 867–867 (1975).

10. M. C. Jonikas, S. R. Collins, V. Denic, E. Oh, E. M. Quan, V. Schmid, J. Weibezahn, B. Schwappach, P. Walter, J. S. Weissman, M. Schuldiner, Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science (80-.).* **323**, 1693–1697 (2009).

11. M. C. S. Lee, E. A. Miller, J. Goldberg, L. Orci, R. Schekman, Bi-directional protein transport between the ER and Golgi. *Annu. Rev. Cell Dev. Biol.* **20**, 87–123 (2004).

12. Y. Guo, D. W. Sirkis, R. Schekman, Protein sorting at the trans-Golgi network. *Annu. Rev. Cell Dev. Biol.* **30**, 169–206 (2014).

13. R. Schekman, P. Novick, 23 Genes, 23 Years Later. *Cell* **116**, 13–15 (2004).

14. A. Feizi, T. Österlund, D. Petranovic, S. Bordel, J. Nielsen, Genome-Scale Modeling of the Protein Secretory Machinery in Yeast. *PLoS One* **8** (2013).

15. P. J. Rapiejko, R. Gilmore, Empty site forms of the SRP54 and SRα GTPase mediate targeting of ribosome-nascent chain complexes to the endoplasmic reticulum. *Cell* **89**,

703–713 (1997).

16. B. Van Den Berg, W. M. C. Jr, I. Collinson, Y. Modis, E. Hartmann, S. C. Harrison, T. A. Rapoport, X-ray structure of a protein-conducting channel. **427** (2004).

17. E. Bause, Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **209**, 331–336 (1983).

18. S. Strahl-Bolsinger, M. Gentzsch, W. Tanner, Protein O-mannosylation. *Biochim. Biophys. Acta - Gen. Subj.* **1426**, 297–307 (1999).

19. L. Ellgaard, M. Molinari, A. Helenius, Setting the standards: Quality control in the secretory pathway. *Science (80-.).* **286**, 1882–1888 (1999).

20. C. Xu, D. T. W. Ng, Glycosylation-directed quality control of protein folding. *Nat. Rev. Mol. Cell Biol.* **16**, 742–752 (2015).

21. M. Delic, M. Valli, A. B. Graf, M. Pfeffer, D. Mattanovich, B. Gasser, The secretory pathway: Exploring yeast diversity. *FEMS Microbiol. Rev.* **37**, 872–914 (2013).

22. B. Meusser, C. Hirsch, E. Jarosch, T. Sommer, ERAD: The long road to destruction. *Nat. Cell Biol.* **7**, 766–772 (2005).

23. T. K. Chaudhuri, S. Paul, Protein-misfolding diseases and chaperone-based therapeutic approaches. *FEBS J.* **273**, 1331–1349 (2006).

24. J. Winderickx, C. Delay, A. De Vos, H. Klinger, K. Pellens, T. Vanhelmont, F. Van Leuven, P. Zabrocki, Protein folding diseases and neurodegeneration: Lessons learned from yeast. *Biochim. Biophys. Acta - Mol. Cell Res.* (2008) https:/doi.org/10.1016/j.bbamcr.2008.01.020.

25. C. Patil, P. Walter, Intracellular signaling from the endoplasmic reticulum to the nucleus : the unfolded protein response in yeast and mammals. *Curr. Opin. Cell Biol.*, 349–356 (2001).

26. K. J. Travers, C. K. Patil, L. Wodicka, D. J. Lockhart, J. S. Weissman, P. Walter, S. Francisco, Functional and Genomic Analyses Reveal an Essential Coordination between the Unfolded Protein Response and ER-Associated Degradation. *Cell* **101**, 249–258 (2000).

27. Y. Kimata, Y. Ishiwata-kimata, S. Yamada, K. Kohno, Yeast unfolded protein response pathway regulates expression of genes for anti-oxidative stress and for cell surface proteins. *Genes to Cells*, 59–69 (2006).

28. M. Saloheimo, M. Valkonen, M. Penttilä, Activation mechanisms of the HACI-mediated unfolded protein response in filamentous fungi. *Mol. Microbiol.* **47**, 1149–1161 (2003).

29. J. Dancourt, C. Barlowe, Protein sorting receptors in the early secretory pathway. *Annu. Rev. Biochem.* **79**, 777–802 (2010).

30. P. P. Poon, D. Cassel, A. Spang, M. Rotman, E. Pick, R. A. Singer, G. C. Johnston, Retrograde transport from the yeast Golgi is mediated by two ARF GAP proteins with overlapping function. *EMBO J.* **18**, 555–564 (1999).

31. H. R. B. Pelham, Sorting and retrieval between the endoplasmic reticulum and Golgi apparatus. *Curr. Opin. Cell Biol.* **7**, 530–535 (1995).

32. H. Kitano, Systems biology: A brief overview. *Science (80-.).* **295**, 1662–1664 (2002).

33. T. Ideker, O. Ozier, B. Schwikowski, F. Andrew, Molecular Interaction Networks. *Encycl. Syst. Biol.* **18**, 1453–1453 (2013).

34. K. R. Patil, J. Nielsen, Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2685–2689 (2005).

35. A. P. Oliveira, K. R. Patil, J. Nielsen, Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Syst. Biol.* **2**, 1–16 (2008).

36. J. Almquist, M. Cvijovic, V. Hatzimanikatis, J. Nielsen, M. Jirstrand, Kinetic models in industrial biotechnology - Improving cell factory performance. *Metab. Eng.* **24**, 38–60 (2014).

37. A. Bordbar, J. M. Monk, Z. A. King, B. O. Palsson, Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* **15**, 107–120 (2014).

38. H. Lu, F. Li, B. J. Sánchez, Z. Zhu, G. Li, I. Domenzain, S. Marcišauskas, P. M. Anton, D. Lappa, C. Lieven, M. E. Beber, N. Sonnenschein, E. J. Kerkhoven, J. Nielsen, A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* **10** (2019).

39. G. Walsh, Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **36**, 1136–1145 (2018).

40. N. Ferrer-Miralles, J. Domingo-Espín, J. Corchero, E. Vázquez, A. Villaverde, Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* **8**, 1–8 (2009).

41. J. L. Martínez, L. Liu, D. Petranovic, J. Nielsen, Pharmaceutical protein production by yeast: Towards production of human blood proteins by microbial fermentation. *Curr. Opin. Biotechnol.* **23**, 965–971 (2012).

42. G. Villadsen, J., Nielsen, J., & Lidén, *Bioreaction Engineering Principles* (Springer Science & Business Media, 2011).

43. I. Diers, E. Rasmussen, P. Larsen, I. Kjaersig, Yeast fermentation processes for insulin production. *Bioprocess Technol.* **13**, 166–176 (1991).

44. Z. Liu, K. E. J. Tyo, J. L. Martínez, D. Petranovic, J. Nielsen, Different expression systems for production of recombinant proteins in Saccharomyces cerevisiae. *Biotechnol. Bioeng.* **109**, 1259–1268 (2012).

45. J. P. van Dijken, W. A. Scheffers, Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiol. Lett.* **32**, 199–224 (1986).

46. K. Campbell, J. Vowinckel, M. A. Keller, M. Ralser, Methionine Metabolism Alters Oxidative Stress Resistance via the Pentose Phosphate Pathway. *Antioxidants Redox Signal.* **24**, 543–547 (2016).

47. D. W. Raiford, E. M. Heizer, R. V. Miller, H. Akashi, M. L. Raymer, D. E. Krane, Do amino acid biosynthetic costs constrain protein evolution in Saccharomyces cerevisiae? *J. Mol. Evol.* **67**, 621–630 (2008).

48. S. S. Dwight, Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Res.* **30**, 69–72 (2002).

49. F. Di, C. Malina, K. Campbell, M. Mormino, J. Fuchs, Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. *Proc. Natl. Acad. Sci.* (2020) https:/doi.org/10.1073/pnas.1918216117.

50. R. J. Haselbeck, L. McAlister-Henn, Function and expression of yeast mitochondrial NAD- and NADP-specific isocitrate dehydrogenases. *J. Biol. Chem.* **268**, 12116–12122 (1993).

51. A. G. Hinnebusch, Translational regulation of GCN4 and the general amino acid control of yeast. *Annu. Rev. Microbiol.* **59**, 407–450 (2005).

52. M. Garcia-Barrio, J. Dong, V. A. Cherkasova, X. Zhang, F. Zhang, S. Ufano, R. Lai, J. Qin, A. G. Hinnebusch, Serine 577 is phosphorylated and negatively affects the tRNA binding and eIF2α kinase activities of GCN2. *J. Biol. Chem.* **277**, 30675–30683 (2002).

53. K. Natarajan, M. R. Meyer, B. M. Jackson, D. Slade, C. Roberts, A. G. Hinnebusch, M. J. Marton, Transcriptional Profiling Shows that Gcn4p Is a Master Regulator of Gene Expression during Amino Acid Starvation in Yeast. *Mol. Cell. Biol.* **21**, 4347–4368 (2001).

54. S. A. Wek, S. Zhu, R. C. Wek, The histidyl-tRNA synthetase-related sequence in the eIF-2 alpha protein kinase GCN2 interacts with tRNA and is required for activation in response to starvation for different amino acids. *Mol. Cell. Biol.* **15**, 4497–4506 (1995).

55. W. Yuan, S. Guo, J. Gao, M. Zhong, G. Yan, W. Wu, Y. Chao, Y. Jiang, General control nonderepressible 2 (GCN2) kinase inhibits target of rapamycin complex 1 in response to amino acid starvation in saccharomyces cerevisiae. *J. Biol. Chem.* **292**, 2660–2669 (2017).

56. A. Soulard, A. Cremonesi, S. Moes, F. Schütz, P. Jenö, M. N. Hall, The rapamycin-sensitive phosphoproteome reveals that TOR controls protein kinase A toward some but not all substrates. *Mol. Biol. Cell* **21**, 3475–3486 (2010).

57. Z. Liu, J. Hou, J. L. Martínez, D. Petranovic, J. Nielsen, Correlation of cell growth and heterologous protein production by Saccharomyces cerevisiae. *Appl. Microbiol. Biotechnol.* **97**, 8955–8962 (2013).

58. M. Huang, J. Bao, B. M. Hallström, D. Petranovic, J. Nielsen, Efficient protein production by yeast requires global tuning of metabolism. *Nat. Commun.* **8** (2017).

59. T. Klein, J. Niklas, E. Heinzle, Engineering the supply chain for protein production/secretion in yeasts and mammalian cells. *J. Ind. Microbiol. Biotechnol.* **42**, 453–464 (2015).

60. W. A. Wilson, P. J. Roach, M. Montero, E. Baroja-Fernández, F. J. Muñoz, G. Eydallin, A. M. Viale, J. Pozueta-Romero, Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiol. Rev.* **34**, 952–985 (2010).

61. M. A. Singer, S. Lindquist, Multiple effects of trehalose on protein folding *invitro* and *invivo*. *Mol. Cell* **1**, 639–48 (1998).

62. E. O. Voit, Biochemical and genomic regulation of the trehalose cycle in yeast: Review of observations and canonical model analysis. *J. Theor. Biol.* **223**, 55–78 (2003).

63. G. Wang, S. M. Björk, M. Huang, Q. Liu, K. Campbell, J. Nielsen, H. N. Joensson, D. Petranovic, RNAi expression tuning, microfluidic screening, and genome recombineering for improved protein production in Saccharomyces cerevisiae . *Proc. Natl. Acad. Sci.* **116**, 9324–9332 (2019).

64. J. Bao, M. Huang, D. Petranovic, J. Nielsen, Moderate expression of SEC16 increases protein secretion by Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **83**, 1–15 (2017).

65. M. Delic, M. Valli, A. B. Graf, M. Pfeffer, D. Mattanovich, B. Gasser, The secretory pathway: Exploring yeast diversity. *FEMS Microbiol. Rev.* **37**, 872–914 (2013).

66. U. Von Stockar, J. S. Liu, Does microbial life always feed on negative entropy? Thermodynamic analysis of microbial growth. *Biochim. Biophys. Acta - Bioenerg.* **1412**, 191–211 (1999).

67. P. Van Hoek, J. P. Van Dijken, J. T. Pronk, Effect of specific growth rate on fermentative capacity of baker's yeast. *Appl. Environ. Microbiol.* **64**, 4226–4233 (1998).

68. P. J. Lahtvee, B. J. Sánchez, A. Smialowska, S. Kasvandik, I. E. Elsemman, F. Gatto, J. Nielsen, Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. *Cell Syst.* **4**, 495-504.e5 (2017).

69. M. Mori, S. Schink, D. W. Erickson, U. Gerland, T. Hwa, Quantifying the benefit of a proteome reserve in fluctuating environments. *Nat. Commun.* **8**, 1–8 (2017).

70. S. Klumpp, M. Scott, S. Pedersen, T. Hwa, Molecular crowding limits translation and cell growth. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16754–16759 (2013).

71. E. Metzl-Raz, M. Kafri, G. Yaakov, I. Soifer, Y. Gurvich, N. Barkai, Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *Elife* **6**, 1–21 (2017).

72. Z. Shi, K. Fujii, K. M. Kovary, N. R. Genuth, H. L. Röst, M. N. Teruel, M. Barna, Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol. Cell* **67**, 71-83.e7 (2017).

73. M. B. Ferretti, J. L. Barre, K. Karbstein, Translational Reprogramming Provides a Blueprint for Cellular Adaptation. *Cell Chem. Biol.* **25**, 1372-1379.e3 (2018).

74. N. Segev, J. E. Gerst, Specialized ribosomes and specific ribosomal protein paralogs control translation of mitochondrial proteins. *J. Cell Biol.* **217**, 1155–1155 (2018).

75. N. Slavov, S. Semrau, E. Airoldi, B. Budnik, A. van Oudenaarden, Differential Stoichiometry among Core Ribosomal Proteins. *Cell Rep.* **13**, 865–873 (2015).

76. J. De La Cruz, K. Karbstein, J. L. Woolford, Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annu. Rev. Biochem.* **84**, 93–129 (2015).

77. P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D. J. Pappin, Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004).

78. J. R. Warner, The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **0004**, 437–440 (1999).

79. M. K. Thompson, M. F. Rojas-Duran, P. Gangaramani, W. V. Gilbert, The ribosomal protein Asc1/RACK1 is required for efficient translation of short mRNAs. *Elife* **5**, 1–22 (2016).

80. A. Des Georges, V. Dhote, L. Kuhn, C. U. T. Hellen, T. V. Pestova, J. Frank, Y. Hashem, Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature* **525**, 491–495 (2015).

81. N. Opitz, K. Schmitt, V. Hofer-Pretz, B. Neumann, H. Krebber, G. H. Braus, O. Valerius, Capturing the Asc1p/Receptor for Activated C Kinase 1 (RACK1) Microenvironment at the head region of the 40s ribosome with quantitative BioID in Yeast. *Mol. Cell. Proteomics* **16**, 2199–2218 (2017).

82. R. A. Crawford, G. D. Pavitt, Translational regulation in response to stress in Saccharomyces cerevisiae. *Yeast* **36**, 5–21 (2019).

83. E. J. O'Brien, J. A. Lerman, R. L. Chang, D. R. Hyduke, B. Ø. Palsson, Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* **9**, 693 (2013).

84. I. E. Elsemman, A. R. Prado, P. Grigaitis, M. G. Albornoz, V. Harman, S. Holman, J. van Heerden, F. J. Bruggeman, M. M. M. Bisschops, N. Sonnenschein, S. Hubbard, R. Beynon, P. Daran-Lapujade, J. Nielsen, B. Teusink, Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *bioRxiv*, 2021.06.11.448029 (2021).

85. L. Qi, B. Tsai, P. Arvan, New Insights into the Physiological Role of Endoplasmic Reticulum-Associated Degradation. *Trends Cell Biol.* **27**, 430–440 (2017).

86. S.-B. Qian, M. F. Princiotta, J. R. Bennink, J. W. Yewdell, Characterization of rapidly degraded polypeptides in mammalian cells reveals a novel layer of nascent protein quality control. *J. Biol. Chem.* **281**, 392–400 (2006).

87. C. C. Glembotski, Endoplasmic reticulum stress in the heart. *Circ. Res.* **101**, 975–984 (2007).

88. A. Stolz, D. H. Wolf, Use of CPY and its derivatives to study protein quality control in various cell compartments. *Methods Mol. Biol.* **832**, 489–504 (2012).

89. C. M. Haynes, E. A. Titus, A. A. Cooper, Degradation of misfolded proteins prevents ER-derived oxidative stress and cell death. *Mol. Cell* **15**, 767–776 (2004).

90. K. A. Geiler-Samerotte, M. F. Dion, B. A. Budnik, S. M. Wang, D. L. Hartl, D. A. Drummond, Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 680–685 (2011).

91. H. S. Choi, S. Y. Lee, T. Y. Kim, H. M. Woo, In silico identification of gene amplification targets for improvement of lycopene production. *Appl. Environ. Microbiol.* **76**, 3097–3105 (2010).

92. Q. Qi, F. Li, R. Yu, M. K. M. Engqvist, V. Siewers, J. Fuchs, J. Nielsen, Different Routes of Protein Folding Contribute to Improved Protein Production in

Saccharomyces cerevisiae. *MBio* **11**, e02743-20 (2020).

93. H. Tang, X. Bao, Y. Shen, M. Song, S. Wang, C. Wang, J. Hou, Engineering protein folding and translocation improves heterologous protein secretion in Saccharomyces cerevisiae. *Biotechnol. Bioeng.* **112**, 1872–1882 (2015).

94. B. C. Hann, C. J. Stirling, P. Walter, SEC65 gene product is a subunit of the yeast signal recognition particle required for its integrity. *Nature* **356**, 532–533 (1992).

95. T. Lodi, B. Neglia, C. Donnini, Secretion of human serum albumin by Kluyveromyces lactis overexpressing KlPDI1 and KlERO1. *Appl. Environ. Microbiol.* **71**, 4359–4363 (2005).

96. A. E. Wentz, E. V Shusta, A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl. Environ. Microbiol.* **73**, 1189–1198 (2007).

97. T. Yu, Y. J. Zhou, M. Huang, Q. Liu, R. Pereira, F. David, J. Nielsen, Reprogramming Yeast Metabolism from Alcoholic Fermentation to Lipogenesis. *Cell* **174**, 1549-1558.e14 (2018).

98. Y. Chen, J. Nielsen, Energy metabolism controls phenotypes by protein efficiency and allocation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 17592–17597 (2019).

99. J. Nocon, M. G. Steiger, M. Pfeffer, S. B. Sohn, T. Y. Kim, M. Maurer, H. Rußmayer, S. Pflügl, M. Ask, C. Haberhauer-Troyer, K. Ortmayr, S. Hann, G. Koellensperger, B. Gasser, S. Y. Lee, D. Mattanovich, Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. *Metab. Eng.* **24**, 129–138 (2014).

100. J. Ptacek, *et al.*, Global analysis of protein phosphorylation in yeast. *Nature* **438**, 679–684 (2005).

101. J. Zhang, S. Vaga, P. Chumnanpuen, R. Kumar, G. N. Vemuri, R. Aebersold, J. Nielsen, Mapping the interaction of Snf1 with TORC1 in Saccharomyces cerevisiae. *Mol. Syst. Biol.* **7**, 1–11 (2011).

102. B. Gasser, M. Maurer, J. Gach, R. Kunert, D. Mattanovich, Engineering of Pichia pastoris for improved production of antibody fragments. *Biotechnol. Bioeng.* **94**, 353–361 (2006).

103. H. Tang, X. Bao, Y. Shen, M. Song, S. Wang, C. Wang, J. Hou, Engineering protein folding and translocation improves heterologous protein secretion in Saccharomyces cerevisiae. *Biotechnol. Bioeng.* **112**, 1872–1882 (2015).

104. P. B. Besada-Lombana, N. A. Da Silva, Engineering the early secretory pathway for increased protein secretion in Saccharomyces cerevisiae. *Metab. Eng.* **55**, 142–151 (2019).

105. S. Hui, J. M. Silverman, S. S. Chen, D. W. Erickson, M. Basan, J. Wang, T. Hwa, J. R. Williamson, Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11**, 784 (2015).

106. J. D. Dinman, Pathways to Specialized Ribosomes: The Brussels Lecture. *J. Mol. Biol.* **428**, 2186–2194 (2016).

107. S. Komili, N. G. Farny, F. P. Roth, P. A. Silver, Functional Specificity among Ribosomal Proteins Regulates Gene Expression. *Cell* **131**, 557–571 (2007).

108. E. Emmott, M. Jovanovic, N. Slavov, Ribosome Stoichiometry: From Form to Function. *Trends Biochem. Sci.* **44**, 95–109 (2019).